Health Technology Assessment 2008; Vol. 12: No. 24

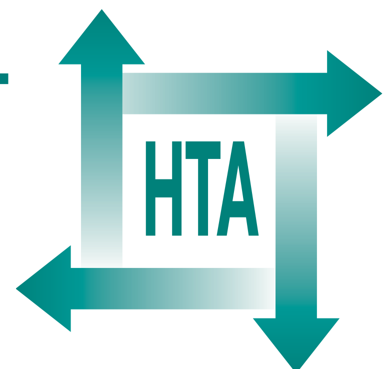
Appendices

Go to main text

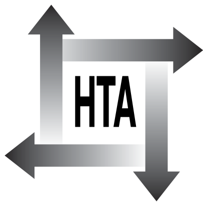
A review and critical appraisal of measures of therapist–patient interactions in mental health settings

J Cahill, M Barkham, G Hardy, S Gilbody, D Richards, P Bower, K Audin and J Connell

June 2008



Health Technology Assessment
NHS R&D HTA Programme
www.hta.ac.uk



How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:
HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Appendix 1

Search strategy used for scoping review (1886–2002)

1. exp Therapeutic Processes/
2. exp Professional Consultation/
3. exp Client Attitudes/
4. exp Health Personnel Attitudes/
5. psychologist attitudes/ or therapist attitudes/
6. counselor attitudes/
7. interpersonal interaction/
8. interpersonal communication/
9. client satisfaction/
10. exp "TRUST (SOCIAL BEHAVIOR)"/
11. ((clinician$ or professional$ or physician$ or doctor$ or nurse$ or therapist$ or psychiatrist$ or psychologist$ or counsel$) adj2 (client$ or patient$) adj2 (interaction$ or relation$ or communicat$)).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
12. "3430".cc.
13. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
14. therapeutic bond$.mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
15. concordance$.mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
16. core skill$.mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
17. (patient$ adj2 (centr$ or center$) adj2 interview$).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
18. (patient$ adj2 (centr$ or center$)).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
19. (expectation$ adj2 outcome$).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
20. exp Countertransference/ or counter transference.mp.
21. working relationship$.mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
22. (mutual adj investment adj company).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
23. boundaries.mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
24. ((adher$ or compliance or comply or complies) adj4 treatment).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
25. (sharing adj2 power).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
26. (turn$ adj3 treatment).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
27. trust.mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
28. therapeutic process$.mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
29. (keep$ adj1 touch adj2 service$).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
30. (relationship adj2 quality).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
31. client participation/
32. exp Decision Making/
33. treatment planning/
34. treatment compliance/
35. exp professional consultation/
36. (client$ attitude$ or psychologist$ attitude$ or psychiatrist$ attitude$ or therapist$ attitude$).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
37. ((clinician$ or professional$ or physician$ or doctor$ or nurse$ or therapist$ or psychiatrist$ or psychologist$ or counsel$ or patient$ or client$ or parent$ or carer$ or inpatient$ or outpatient$) adj4 (friendship$ or transference or relationship$ or agreement$ or communication$ or dialogue$ or contract or involvement or trust or engagement or disagreement or power or alliance$ or partnership$ or attachment$ or interaction$ or consultation$)).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
38. 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37
39. 13 or 38

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

40. ((systematic or literature) adj review).mp. [mp=title, abstract, heading word, table of contents, key phrase identifiers]
41. 39 and 40
42. limit 39 to "0700 editorials"
43. limit 39 to "1300 literature review/research review"
44. limit 39 to 5000 collected works
45. 41 or 42 or 43 or 44
46. limit 45 to english language

Appendix 2

Data summary sheet used for scoping review

- Title/authors
- Method of review/timespan
- Topic/themes under review
- Theoretical orientations/psychotherapies covered
- Details of setting/therapist
- Population
- Therapist–patient interactions (list)
- Measures used (list)
- Brief summary of findings
- Other reviews cited
- Key authors

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Appendix 3

Measures search strategy

PsycINFO (1886–2002)

1. exp Therapeutic Processes/
2. exp Professional Consultation/
3. exp Client Attitudes/
4. exp Health Personnel Attitudes/
5. psychologist attitudes/ or therapist attitudes/
6. counselor attitudes/
7. interpersonal interaction/
8. interpersonal communication/
9. client satisfaction/
10. exp "TRUST (SOCIAL BEHAVIOR)"/
11. ((clinician$ or professional$ or physician$ or doctor$ or nurse$ or therapist$ or psychiatrist$ or psychologist$ or counsel$) adj2 (client$ or patient$) adj2 (interaction$ or relation$ or communicat$)).mp. [mp=title, abstract, heading word, table of contents, key concepts]
12. "3430".cc.
13. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
14. therapeutic bond$.mp. [mp=title, abstract, heading word, table of contents, key concepts]
15. concordance$.mp. [mp=title, abstract, heading word, table of contents, key concepts]
16. core skill$.mp. [mp=title, abstract, heading word, table of contents, key concepts]
17. (patient$ adj2 (centr$ or center$) adj2 interview$).mp. [mp=title, abstract, heading word, table of contents, key concepts]
18. (patient$ adj2 (centr$ or center$)).mp. [mp=title, abstract, heading word, table of contents, key concepts]
19. (expectation$ adj2 outcome$).mp. [mp=title, abstract, heading word, table of contents, key concepts]
20. exp Countertransference/ or counter transference.mp.
21. working relationship$.mp. [mp=title, abstract, heading word, table of contents, key concepts]
22. (mutual adj investment adj company).mp. [mp=title, abstract, heading word, table of contents, key concepts]
23. boundaries.mp. [mp=title, abstract, heading word, table of contents, key concepts]
24. ((adher$ or compliance or comply or complies) adj4 treatment).mp. [mp=title, abstract, heading word, table of contents, key concepts]
25. (sharing adj2 power).mp. [mp=title, abstract, heading word, table of contents, key concepts]
26. (turn$ adj3 treatment).mp. [mp=title, abstract, heading word, table of contents, key concepts]
27. trust.mp. [mp=title, abstract, heading word, table of contents, key concepts]
28. therapeutic process$.mp. [mp=title, abstract, heading word, table of contents, key concepts]
29. (keep$ adj1 touch adj2 service$).mp. [mp=title, abstract, heading word, table of contents, key concepts]
30. (relationship adj2 quality).mp. [mp=title, abstract, heading word, table of contents, key concepts]
31. client participation/
32. exp Decision Making/
33. treatment planning/
34. treatment compliance/
35. exp professional consultation/
36. (client$ attitude$ or psychologist$ attitude$ or psychiatrist$ attitude$ or therapist$ attitude$).mp. [mp=title, abstract, heading word, table of contents, key concepts]
37. ((clinician$ or professional$ or physician$ or doctor$ or nurse$ or therapist$ or psychiatrist$ or psychologist$ or counsel$ or patient$ or client$ or parent$ or carer$ or inpatient$ or outpatient$) adj4 (friendship$ or transference or relationship$ or agreement$ or communication$ or dialogue$ or contract or involvement or trust or engagement or disagreement or power or alliance$ or partnership$ or attachment$ or interaction$ or consultation$)).mp. [mp=title, abstract, heading word, table of contents, key concepts]
38. 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37
39. 13 or 38
40. exp PSYCHOMETRICS/
41. psychometri$.mp.
42. exp Rating Scales/
43. rating scal$.mp.
44. exp QUESTIONNAIRES/
45. questionaire$.mp.
46. questionnaire$.mp.
47. "Personality Scales & Inventories ".cc.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

48. exp Test Construction/
49. exp Test Validity/
50. or/40-49
51. 39 and 50
52. limit 51 to english language
53. or/40-43,47-49
54. 39 and 53
55. limit 54 to english language

MEDLINE (1886–2002)

1. exp psychotherapeutic processes/
2. Attitude of Health Personnel/
3. exp professional-patient relations/
4. exp interpersonal relations/
5. exp consumer satisfaction/
6. ((clinician$ or professional$ or physician$ or doctor$ or nurse$ or therapist$ or psychiatrist$ or psychologist$ or counsel$) adj2 (client$ or patient$) adj2 (interaction$ or relations$ or communicat$)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
7. therapeutic bond$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
8. concordance$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
9. core skill$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
10. (patient$ adj2 (centr$ or center$)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
11. (expectation$ adj2 outcome$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
12. "Countertransference (Psychology)"/ or countertransference.mp.
13. working relationship$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
14. (mutual adj investment adj company).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
15. boundaries.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
16. ((adher$ or compliance or comply or complies) adj4 treatment).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
17. (sharing adj2 power).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
18. (turn$ adj3 treatment$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
19. trust.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
20. therapeutic process$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
21. (keeps adj1 touch adj2 service$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
22. (relationship adj2 quality).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
23. Patient Participation/
24. exp Decision Making/
25. Patient Compliance/
26. (client$ attitude$ or psychologist$ attitude$ or psychiatrist$ attitude$ or therapist$ attitude$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
27. ((clinician$ or professional$ or physician$ or doctor$ or nurse$ or therapist$ or psychiatrist$ or psychologist$ or counsel$ or patient$ or client$ or parent$ or carer$ or inpatient$ or outpatient$) adj4 (friendship$ or transference$ or relationship$ or agreement$ or communication$ or dialogue$ or contract or involvement or trust or engagement or disagreement or power or alliance$ or partnership$ or attachment$ or interaction$ or consultation$)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
28. or/1-27
29. limit 28 to (editorial or review or review, academic or review, multicase or review, tutorial or review literature)
30. exp mental health/
31. exp psychiatry/
32. exp psychology/
33. psychiat$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
34. mental$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
35. psychol$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
36. or/30-35
37. 29 and 36
38. Psychometrics/
39. psychometric$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

40. instrument$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
41. exp Psychiatric Status Rating Scales/
42. rating scale$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
43. Questionnaires/
44. questionaire$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
45. questionnaire$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
46. Evaluation Studies/
47. evaluation stud$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]
48. or/38-47
49. 28 and 48
50. 37 and 48

EMBASE (1886–2002)

1. exp Doctor Patient Relation/
2. exp Physician Attitude/
3. exp Nurse Attitude/
4. exp Patient Attitude/
5. exp Nurse Patient Relationship/
6. exp Human Relation/
7. exp Attitude/ and exp Health Care Personnel/
8. exp Interpersonal Communication/
9. exp Patient Satisfaction/
10. trust.mp.
11. ((clinician$ or professional$ or physician$ or doctor$ or nurse$ or therapist$ or psychiatrist$ or psychologist$ or counsellor$ or counselor$) adj2 (client$ or patient$) adj2 (interaction$ or relation$ or communicat$ or therapeutic process$)).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
12. (nhs adj trust).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
13. 10 not 12
14. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 11 or 13
15. therapeutic bond$.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
16. concordance$.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
17. core skill$.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
18. (patient$ adj (center$ or centre$) adj interview$).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
19. (expectation$ adj2 outcome$).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
20. (counter adj2 transference).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
21. exp Counter Transference/
22. (working adj relationship$).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
23. (mutual adj investment adj company).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
24. ((adher$ or comply$ or compliance or complies$) adj4 treatment$).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
25. (sharing adj2 power).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
26. (turn$ adj3 treatment).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
27. therapeutic process$.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
28. (keep$ adj2 touch adj2 service$).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
29. (relationship$ adj2 quality).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
30. ((clinician$ or professional$ or physician$ or nurse$ or doctor$ or therapist$ or psychiatrist$ or psychologist$ or counselor$ or counsellor$ or patient$ or client$ or

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

parent$ or carer$ or inpatient$ or outpatient$) adj4 (friendship$ or transference or relations$ or agreement or communicat$ or dialogue$ or contract or involvement or trust or engagement or disagreement$ or power or alliance or attachment$ or consultation$ or partner$)).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]

31. client participation.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
32. exp Decision Making/
33. exp Treatment Planning/
34. exp Patient Compliance/
35. professional consultat$.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
36. client attitude$.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
37. psychologist$ attitude$.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
38. counsel?or attitude$.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
39. (interpersonal adj2 interaction$).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
40. client satisfaction.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
41. 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40
42. 14 or 41
43. limit 42 to review
44. limit 42 to editorial
45. ((literature or systematic) adj2 review$).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
46. 42 and 45
47. 46 or 43 or 44
48. limit 42 to english language
49. exp Mental health/
50. exp Psychiatry/
51. exp Psychiatrist/
52. exp psychology/
53. exp psychologist/
54. exp Mental Disease/
55. 49 or 50 or 51 or 52 or 53 or 54
56. (psychol$ or psychiat$).mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
57. 55 or 56
58. 48 and 57
59. exp Psychometry/
60. psychomet$.mp.
61. instrument$.mp.
62. Rating Scale/
63. rating scale$.mp.
64. exp Questionnaire/
65. questionaire$.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
66. questionnaire$.mp. [mp=title, abstract, subject headings, drug trade name, original title, device manufacturer, drug manufacturer name]
67. evaluation/
68. evaluation stud$.mp.
69. or/59-68
70. 58 and 69

CINAHL (1886–2002)

1. exp psychotherapy/
2. exp Attitude of Health Personnel/
3. exp professional-patient relations/
4. exp interpersonal relations/
5. exp consumer satisfaction/
6. ((clinician$ or professional$ or physician$ or doctor$ or nurse$ or therapist$ or psychiatrist$ or psychologist$ or counsel$) adj2 (client$ or patient$) adj2 (interaction$ or relations$ or communicat$)).mp. [mp=title, cinahl subject heading, abstract, instrumentation]
7. therapeutic bond$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
8. concordance$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
9. core skill$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
10. (patient$ adj2 (centr$ or center$)).mp. [mp=title, cinahl subject heading, abstract, instrumentation]
11. (expectation$ adj2 outcome$).mp. [mp=title, cinahl subject heading, abstract, instrumentation]

12. "Countertransference (Psychology)"/ or countertransference.mp.
13. working relationship$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
14. (mutual adj investment adj company).mp. [mp=title, cinahl subject heading, abstract, instrumentation]
15. boundaries.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
16. ((adher$ or compliance or comply or complies) adj4 treatment).mp. [mp=title, cinahl subject heading, abstract, instrumentation]
17. (sharing adj2 power).mp. [mp=title, cinahl subject heading, abstract, instrumentation]
18. (turn$ adj3 treatment$).mp. [mp=title, cinahl subject heading, abstract, instrumentation]
19. trust.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
20. therapeutic process$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
21. (keeps adj1 touch adj2 service$).mp. [mp=title, cinahl subject heading, abstract, instrumentation]
22. (relationship adj2 quality).mp. [mp=title, cinahl subject heading, abstract, instrumentation]
23. Patient Participation/
24. exp Decision Making/
25. Patient Compliance/
26. (client$ attitude$ or psychologist$ attitude$ or psychiatrist$ attitude$ or therapist$ attitude$).mp. [mp=title, cinahl subject heading, abstract, instrumentation]
27. ((clinician$ or professional$ or physician$ or doctor$ or nurse$ or therapist$ or psychiatrist$ or psychologist$ or counsel$ or patient$ or client$ or parent$ or carer$ or inpatient$ or outpatient$) adj4 (friendship$ or transference$ or relationship$ or agreement$ or communication$ or dialogue$ or contract or involvement or trust or engagement or disagreement or power or alliance$ or partnership$ or attachment$ or interaction$ or consultation$)).mp. [mp=title, cinahl subject heading, abstract, instrumentation]
28. or/1-27
29. exp mental health/
30. exp psychiatry/
31. exp psychology/
32. psychiat$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
33. mental$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
34. psychol$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
35. or/29-34
36. Psychometrics/
37. psychometric$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
38. instrument$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
39. exp Research Instruments/
40. rating scale$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
41. Questionnaires/
42. questionaire$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
43. questionnaire$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
44. Evaluation Studies/
45. evaluation stud$.mp. [mp=title, cinahl subject heading, abstract, instrumentation]
46. or/36-45
47. 28 and 35
48. 46 and 47

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Appendix 4

Measures excluded by the electronic sieve

16 Client Stimulus Impressions
A client satisfaction form
A client's satisfaction scale
Acceptance Rejection Attitude Scale
Acceptance Rejection Scale
Affect in Play Scale
Analytic instrument measuring inferential communications in psychotherapy
Aspects of therapeutic environments questionnaire
Automatic analysis of speech behaviour in psychotherapy procedure
Bogardus Ethnic Distance Scale
Care Perception Questionnaire
Checklist of Interpersonal Transactions
Client Centered Counseling Progress Record
Client Evaluation of Counselor Scale
Client Expectancy Questionnaire
Client Non-Compliance Code
Client Perception Questionnaire
Client Resistance Code
Client Satisfaction Survey
Client Vocal Quality System
Clients' Personal Reaction Questionnaire
Coding System of Therapist Feedback
Combined Alliance – Short Form
Counseling Interaction Profile
Counseling Proficiency Scale
Counsellor Attitude Scale
Counsellor Reaction Form
Counselor Preference Inventory
Counselor Termination Behaviour Inventory
Current Attachment Relationships Questionnaire
Davis Interpersonal Reactivity Scale
Denver Community Mental Health Questionnaire
Difficult Patient Stress Scale
Facilitative Relationship Indicators Checklist
Family Therapist Alliance Scale
Family Therapist Coding System
First Impression Rating Scale
Free Association Scale
Fundamental Interpersonal Relations Orientation Scale
General Counselor Appeal Questionnaire
Group Environment Scale
Group Therapy Survey – Revised
Heuristic Rating Scales
Initial Homework Non-Compliance Scale
Insession Non-Compliance Scale
Interactional Style Rating Scale
Interpersonal Schema Questionnaire
Krankhetis–Konzpet–Skala
Laing's Interpersonal Perception Method
Measure of Client Perceive Therapist Regard
Measure of Countertransference
Measure of Empathic Accuracy
Measure of Therapist Trustworthiness
Member Leader Scoring System
Microcounseling Skill Discrimination Scale
Munich Patient Satisfaction Scale
Neuropsychological Alliance Scale
Patient Attitude Questionnaire
Perceived Control Scale
Picture Impressions Measure
Psychiatric Care Satisfaction Questionnaire
Psychotherapy Questionnaire
Questionnaire on Procrastination and Estimation of Personal Control
Ratings of Therapist Credibility
Recreational Therapy Rating Form
Relational Empathy Inventory
Resistance Process Rating Scale
Responding to Interpersonal Process Scale
Robert's Apperception Test
Role Expectation and Preference Questionnaire
Satisfaction Index – Mental Health Instrument
Scale for the measurement of empathic understanding
Scale to measure patient collaboration with psychoanalytically oriented therapy
Scaling of communication levels in psychotherapeutic groups
Self-report inventory to assess psychotherapists' styles of interaction with clients
Service Satisfaction Scale
Staff Patient Interaction Response Scale
Stationserfahrungsbogen (SEB) experiences of the inpatient psychotherapeutic process
Stuttgarter Kategorien Inventar
The Consultation Readiness Scale
The Practicum Interaction Observation Form
The Therapist Questionnaire
Therapeutic Relationship Scale
Therapist Alliance Focus Scale
Therapist Client Rating Scale
Therapist Intervention Rating System

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Therapist Personal Reaction Questionnaire
Therapy Attitude Inventory
Therapy Involvement Scale
Two scales measuring openness and awareness in psychotherapy
UKU-ConSat
Understanding Suicide Scale
Utility of repertory grid for measuring treatment process and outcome
Walfish Crisis Contract Scale

Appendix 5

Measures excluded on the basis of content

Archaic Involvement Measure
Body Formation Coding System
California Psychological Inventory
Carers' and Users' Expectations of Services – Users
Empathic Understanding Scale
Charleston Psychiatric Outpatient Satisfaction Scale
Client Satisfaction Questionnaire: Spanish and 18 item
Client Satisfaction Questionnaire – Extended
Client Satisfaction Questionnaire-8: Original and Dutch
Client/Consumer Questionnaire – 23 items
Colorado Client Assessment Record
Compliance Self Rating Scale
Comprehensive Process Analysis
Counseling Evaluation Inventory – French Version
Counselor Rating Form – French Version
Edwards Personal Preference Schedule
Emotional Empathy Scale
Family Therapist Behavior Scale
Family Therapist Rating Scale
Goal Attainment Scaling
Group Therapy Survey
Helping Skills Measure: 12 items
Helping Skills Measure: 13 items
Hogan Empathy Scale
Hospital Relations Technique
Impact Factors Process Scale
Inpatient Consumer Satisfaction Scale
Internal–External Locus of Control of Reinforcement Scale
Intervention Rating Profile
Inventory of Countertransference Behavior
Maslach Burnout Inventory – General Scale
Medical Student Interviewing Performance Questionnaire
Mental Health Locus of Control Scale (Hill)
Mental Health Locus of Control Scale (Wood)
Minnesota Multiphasic Personality Inventory – 2
Multidimensional Adolescent Satisfaction Scale
Nurses Observational Scale for Inpatient Evaluation
Patient Satisfaction Interview
Patient Satisfaction Questionnaire
Process Scoring System
Psychiatric Care Consumer Satisfaction Survey
Psychiatric Care Satisfaction Questionnaire
Psychiatrist's Sphere Of Influence Scale
Quality of Object Relations
Questionnaire for the measurement of psychological reactance
Rutgers Psychotherapy Progress Scale
Satisfaction with Mental Health Care Scale
Satisfaction with Therapy and Therapist Scale
Scale for Counselor Growth Focus
Self Dyadic Perspective Taking Scale and the Other Dyadic Perspective Taking Scale
Service Satisfaction Scale
Verona Service Satisfaction Scale
Verona Service Satisfaction Scale – Intermediate
Verona Service Satisfaction Scale – Short
Sequential Plan Analysis
Structural Analysis of Social Behavior
Suicide Intervention Response Inventory: Original
Suicide Intervention Response Inventory: Revised
Teacher and Pupil Relationship Inventory
Therapeutic Reactance Scale
Treatment Evaluation Inventory
Treatment Evaluation Inventory: 11 items
Treatment Evaluation Inventory: Short
Verbal Report Form
Verona Expectations for Care Scale
Youth Client Satisfaction Questionnaire

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Appendix 6

Database access extraction forms

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Microsoft Access - [FrmMeasure : Form]

File Edit View Insert Format Records Tools Window Help

General | TPI Purpose and Aspects | Utility | Sub Measures | References | Notes

Measure ID: 73

Date Updated: 28/01/03

Updated By: Kanan

Measure: Rutgers Psychotherapy Progress Scale

Acronym: RPPS

Author:

Number of Items: 8

Dimensions/Subscales: No

Administration Method: Questionnaire

ScaleType 1: Ordinal

ScaleType 2: Likert

Country of Publication: USA

Language: English

Publisher:

Copyright:

Cost:

Cost Details:

Manual Available ☑ Details: RPPS and Scoring Manual available from Health and Psychosocial Instruments (HAPI), PO Box 11!

Support Available ☐ Details:

Web/Scanning Options ☐ Details

Switchboard | References | Go to References and Measures | Save Record

Record: 6 of 190

Form View

Start | Draft 2 of database sectio... | Microsoft Access - [F... | Microsoft Office Shortcut Bar

NUM

15:34

FIGURE 4 Page 1 of Measures Form

FIGURE 5 *Page 1 of References Form*

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Appendix 7

Data summary sheet

Measure Title

General details	
Author Language Country of publication/development Publication date Publisher	

Purpose and overview	
<<Introductory text>>	
Theoretical orientation Population details Perspective Measure used by Other versions Notes	

Areas of therapist–patient interaction addressed: Map

Dimensions	
<<Name>> <<Name>>	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
<<Introductory sentence>>	
Split-half Internal consistency Inter-rater Test–retest	

Validity	
<<Introductory sentence>>	
Face Content Criterion (a) concurrent Criterion (b) predictive Construct Convergent Discriminant Factor structure	

Responsiveness	
Discriminative (between individuals) Evaluative (within individual across time)	

Acceptability	
Number of items Administration method Time taken to complete Flesch reading age Translations Access by ethnic minorities	

continued

Feasibility	
Copyright Web or scanning options Training details Administration/process details Support from measure developers FAQ facility	

Precision	
Scale type Normative data	

Notes

Résumé	
Strengths Weaknesses Areas for further research	

Primary references
Secondary references

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Appendix 8

Measure summaries

Measure	ID
Affective Sensitivity Scale (ASS) – Form A	A1
Affective Sensitivity Scale (ASS) – Form C	A2
Affective Sensitivity Scale (ASS) – Form D	A3
Affective Sensitivity Scale (ASS) – Form D-80	A4
Affective Sensitivity Scale (ASS) – Forms E-80 and E-A-2	A5
Kagan Affective Sensitivity Scale (KASS) – Form H	A6
Agnew Relationship Measure (ARM)	A7
Barrett-Lennard Relationship Inventory (BLRI or RI)	B1
California Psychotherapy Alliance Scale (CALPAS)	C1
California Psychotherapy Alliance Scales – Patient (CALPAS-P)	C2
California Psychotherapy Alliance Scales – Rater (CALPAS-R)	C3
California Psychotherapy Alliance Scales – Therapist (CALPAS-T)	C4
California Therapeutic Alliance Rating System (CALTARS)	C5
California Therapeutic Alliance Rating System Scales (CALTARS Scales)	C6
Capacity for Dynamic Process Scale (CDPS)	C7
Carkhuff Scales	C8
Child Psychotherapy Process Scales (CPPS)	C9
Client Attachment to Therapist Scale (CATS)	C10
Client Resistance Scale (CRS)	C11
Coding the Interaction in Psychotherapy (CIP)	C12
Coherence of the Relationship Theme (CRT)	C13
Core Conflictual Relationship Theme (CCRT)	C14
Counseling Evaluation Inventory (CEI)	C15
Counselor Effectiveness Rating Scale (CERS)	C16
Counsellor Effectiveness Scale (CES)	C17
Counselor Rating Form (CRF)	C18
Counselor Rating Form – Short Version (CRF-S)	C19
Counselor Evaluation Rating Scale (CERS)	C20
Counselor Perception Questionnaire (CPQ)	C21
Cross-Cultural Counseling Inventory – Revised (CCCI-R)	C22
Empathy Construct Rating Scale – 23 (ECRS-23)	E1
Empathy Construct Rating Scale – 84 (ECRS-84)	E2
Empathy Test (ET)	E3
Experiencing Scale (EXP)	E4
Family Engagement Questionnaire (FEQ)	F1
Family Therapeutic Alliance Scale (FTAS)	F2
Feminist Self-Disclosure Inventory (FSDI)	F3
Group Assessment of Interpersonal Traits (GAIT)	G1
Helper Behaviour Rating System – Modified Version	H1
Helpful Responses Questionnaire (HRQ)	H2
Helping Alliance Counting Signs Method (HAcs)	H3
Hill Client Verbal Response Category System (HCVRCS)	H4
Hill Interaction Matrix – Form G (HIM-G)	H5
Hill Interaction Matrix – Statement by Statement (HIM-SS)	H6
Hill Counselor Verbal Response Category System (HCVRCS)	H7
Hill Counselor Verbal Response Category System – Revised (HCVRCS-R)	H8
Integrative Psychotherapy Alliance Scale (IPAS)	I1
Intersession Experience Questionnaire (IEQ)	I2
Maslach Burnout Inventory – Therapist and Client Versions (MBI-T and MBI-C)	M1
Missouri Identifying Transference Scale (MITS)	M2
Multicultural Counseling Inventory (MCI)	M3

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Measure	ID
Octant Scale Impact Message Inventory (IMI-C)	O1
Patient Action Scale (PAS)	P1
Penn Helping Alliance Questionnaire (HAq)	P2
Penn Helping Alliance Questionnaire – Revised (HAq-II)	P3
Penn Helping Alliance Rating Scale (HAr)	P4
Psychotherapy Process Inventory (PPI)	P5
Psychotherapy Process Q-Set (PPQS)	P6
Reasons for Ending Treatment Questionnaire (RETQ)	R1
Session Evaluation Questionnaire (SEQ)	S1
Session Evaluation Questionnaire (SEQ) – Form 3	S2
Session Evaluation Questionnaire (SEQ) – Form 4	S3
Session Impacts Scale (SIS)	S4
Therapeutic Alliance Scales for Children	T1
Therapeutic Bond Scales	T2
Therapeutic Factors Inventory (TFI)	T3
Therapist Action Scale (TAS)	T4
Therapist Behavior Scale (TBS)	T5
Therapist Representation Inventory (TRI) – Fourth Section: Record of Dreams	T6
Therapist Representation Inventory (TRI) – Free Response Task	T7
Therapist Representation Inventory (TRI) – Therapist Embodiment Scale (TES)	T8
Therapist Representation Inventory (TRI) – Therapist Involvement Scale (TIS)	T9
Truax and Carkhuff (1967) Scales	T10
Vanderbilt Negative Indicators Scale (VNIS)	V1
Vanderbilt Negative Indicators Scale – Short (VNIS-S)	V2
Vanderbilt Psychotherapy Process Scale – 80 item (VPPS-80)	V3
Vanderbilt Therapeutic Alliance Scale (VTAS)	V4
Working Alliance Inventory – Client (WAI-C)	W1
Working Alliance Inventory – Observer (WAI-O)	W2
Working Alliance Inventory – Therapist (WAI-T)	W3
Working Alliance Inventory – Client – Short (WAI-C-S)	W4
Working Alliance Inventory – Observer – Short (WAI-O-S)	W5
Working Alliance Inventory – Therapist – Short (WAI-T-S)	W6

A1
Affective Sensitivity Scale (ASS) – Form A

General details	
Authors	Campbell RJ, Kagan N, Krathwohl DR
Language	English
Country of publication/development	USA
Publication date	1971
Publisher	NA
Purpose and overview	
Form A is the first stage in the development of the Affective Sensitivity Scale (ASS). The ASS was developed from interpersonal process recall (IPR).[3] The purpose of the ASS is to test a participant's ability to detect and identify the immediate affective state of another (affective sensitivity). After viewing an excerpt of counselling, participants are asked to feel the affective state of the client, and describe it.[1]	
Two kinds of items reflect (a) the client's feelings of him/herself and (b) the relationship between the counsellor and client[1]	
Theoretical orientation	Interpersonal theory
Population details	Non-clinical student counsellors
Perspective	Self-report
Measure used by	Researchers
Other versions	ASS Forms C, D, D-80, E, E-80, E-A-2, F and H
Notes	Form A was completed by 26 student members of a master's degree National Defence Education Act (NDEA) Counselling and Guidance Institute
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: empathy/sensitivity	
Outcome: expression of feelings	
The therapist–client interaction information is derived from a general description of the measure	
Dimensions	
No details	
Reliability	
The split-half reliability of Form A has been partially supported by Kuder–Richardson Formula 20[1]	
Split-half	Kuder–Richardson Formula 20 = 0.57
Internal consistency	39 of the scale's 86 items had significant *t* values
Inter-rater	No details
Test–retest	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Validity	
In assessment of convergent validity, Form A scores were correlated with those of peer-rated counsellor effectiveness obtained at the beginning, and again at the end of the institute. Form A scores correlated negatively with peer ratings from the beginning of the institute and its convergent validity was partially supported (falling just short of adequate) by the correlation between Form A and peer ratings taken at the end of the institute[1]	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	Correlation coefficients between Form A and peer ratings of counsellor effectiveness obtained at the beginning and at the end of the institute were –0.02 and 0.49 respectively
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	Many items were too easy and did not discriminate between high and low scorers[1]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	86
Administration method	Multiple-choice questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1971, *Journal of Counseling Psychology*
Web or scanning options	No details
Training details	No details
Administration/process details	Participants view videotape and kinescope excerpts of actual counselling sessions. They are asked to feel whatever emotions the client felt at the end of the excerpt, and to describe the feeling by checking one of three responses to each item
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Binary
Normative data	No details

continued

Notes	
The article from which this summary was written[1] is based on the first author's doctoral dissertation[2]	
Each of the scale's 86 items was selected from 224 according its ability to discriminate between high and low scorers as measured by total scores on the 224-item scale, and staff and peer ratings of counsellor effectiveness[1]	
Higher scores have been obtained using videotapes, compared to kinescope. This is thought to be due to the better sound quality of the former[1]	
Résumé	
Strengths	Some observations were made from the item analyses, which allowed for certain general observations and the construction of a revised form[1]
Weaknesses	There is only partial support for the split-half reliability of Form A. When correlated with peer ratings of counselling effectiveness, correlation coefficients were inadequate and partial. Also, many items did not discriminate between low and high scorers[1]
Areas for further research	Form A is no longer used in research as, following item analyses, it was replaced by Form B[1]

Primary reference

1. Campbell RJ, Kagan N, Krathwohl DR. The development and validation of a scale to measure affective sensitivity (empathy). *J Counsel Psychol* 1971;**18**:407–12.

Secondary references

2. Campbell RJ. The development and validation of a multiple-choice scale to measure affective sensitivity (empathy). Unpublished doctoral dissertation, Michigan State University; 1967.
3. Kagan N, Krathwohl DR, Farquhar WW. *Interpersonal process recall: stimulated recall by videotape*. East Lansing: Michigan State University, Educational Publication Services; 1965.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

A2
Affective Sensitivity Scale (ASS) – Form C

General details	
Authors	Campbell RJ, Kagan N, Krathwohl DR
Language	English
Country of publication/development	USA
Publication date	1971
Publisher	NA

Purpose and overview

Form C is a version of the ASS. Developed from IPR,[18] the purpose of the ASS is to test the participant's ability to detect and identify the immediate affective state of another (affective sensitivity).[2] The scale has been widely used for training and supervision in the helping professions[7]

After viewing an excerpt of counselling, participants are asked to feel the affective state of the client, and describe it. For each item, the participant chooses one of three responses (one of which correctly identifies the client's feelings, and two are incorrect)[2]

Two kinds of items reflect (a) the client's feelings of him/herself (similar to 'I'm a little confused, I have trouble expressing myself') and (b) the relationship between the offered and client (similar to 'you really understand me, I like that')[2]

Theoretical orientation	Interpersonal theory
Population details	Non-clinical adults and non-clinical adolescents
Perspective	Self-report
Measure used by	Researchers and the helping professions for training and supervision
Other versions	ASS Forms A, D, D-80, E, E-80, E-A-2, F and H
Notes	ASS Form C has been used with: Sample groups consisting of undergraduates, counselling students at master's and doctoral degree levels and experienced training[2] A T-group of 65 participants ranging in age from 17-year-old students to adults in their sixties[3] The form has also been used with teachers[9] The scale has been used with clinical and counselling psychologists in the final phase of doctoral programmes. Therapists emphasised, as their theoretical orientations, self-exploration, insight, current versus past intrapsychic conflict and interpersonal personality theory[6]

Areas of therapist–patient interaction addressed: Map

Therapist engagement: empathy/sensitivity

Outcomes: expression of feelings

The therapist–client interaction information is derived from a general description of the measure

Dimensions

No detail

Reliability

Out of 11 measures of split-half reliability, Kuder–Richardson formula 20 values have been adequate for six, and partially supportive for five[2]

The ASS has demonstrated adequate 1-week test–retest reliability[2]

continued

Split-half	Kuder–Richardson formula 20 = 0.74 (*n* = 232, *p*-value not reported)[2,4] Kuder–Richardson formula 20 values for seven sample groups (three completing Form B twice) ranged from 0.53 to 0.77 (*p*-values not reported)[2]
Internal consistency	No details
Inter-rater	No details
Test–retest	1-week test–retest, $r = 0.75$ ($n = 26$, *p*-values not reported)[2]
Validity	
Support for the construct validity and responsiveness of the ASS has been demonstrated in two studies where one-tailed *t*-tests on pretest–post-test scores upheld hypothesised gains following participation in an offered education programme[2] and groups designed to increase sensitivity[3] The convergent validity of the scale has been assessed with mixed results: Of eight correlations with ratings of sensitivity, as rated by therapists, peers and supervisors, half of the coefficients were adequate, three were partially supportive and one was inadequate. Of seven correlations with offered effectiveness ratings, as rated by supervisors, staff and peers, one was adequate, three were partially supportive and three were inadequate[2] The scale demonstrated convergent validity in a study with teachers that found a significant relationship between the form and observed behaviours[9] The ASS has been correlated with five other measures purporting to tap empathy, and a process measure of client self-exploration, with no significant coefficients found[6] The scale's predictive validity has also been assessed with mixed, though largely unsupportive results: Low scores predicted low peer ratings of offered effectiveness better than high scores predicted high peer ratings of offered effectiveness[2] The scale was correlated with six outcome measures and all resulting coefficients were non-significant and five were negative[6]	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	Low scores on the scale better predicted low peer ratings of counsellor effectiveness, as compared to high scores predicting high peer ratings[2] The scale was correlated with six outcome measures and all resulting coefficients were non-significant and five were negative[6]
Construct	*t*-values for mean pretest–post-test ASS scores were: 2.06 ($p = 0.025$, df = 25), one-tailed test (counsellor education)[2] 2.76 ($p = 0.005$, df = 29), one-tailed test (counsellor education)[2] 2.51 ($p < 0.01$, df = 40), one-tailed test (sensitivity group)[3]
Convergent	A significant relationship has been reported between teachers' observed behaviours and the ASS[9] The ASS has been correlated with five measures purporting to tap empathy, and a process measure of client self-exploration, with no significant coefficients found[6] Spearman (rho) coefficients of Form B scores with: Therapists' rankings of sensitivity were 0.35 ($n = 9$), 0.59 ($n = 9$) and 0.64 ($n = 8$), with an average of 0.53 ($p < 0.01$ for the average, not reported individually)[2,4] Peer rankings of sensitivity were –0.10 ($n = 9$), 0.51 ($n = 9$) and 0.59 ($n = 8$) (*p*-values not reported)[2,4] Supervisors' rankings of sensitivity were 0.32 and 0.38 ($n = 16$, *p*-average = 0.06)[2] Supervisors' rankings of counsellor effectiveness were 0.31 and 0.32 ($n = 16$, average significance at the 0.05 level)[1]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	Correlation coefficients (r_s) of Form B scores ratings of counsellor effectiveness by: Staff were 0.17 ($n = 24$, p not reported), 0.32 ($n = 26$, not significant at the 5% level), and 0.42 ($n = 0.27$, $p < 0.025$)[2] And peers were 0.20 ($n = 24$) and 0.28 ($n = 26$) p-values not reported[2]
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	ASS post-test scores of six T-groups were as assessed for, and demonstrated, sensitivity to differential group experiences. Differences in the pretest–post-test scores of six T-groups ($n = 5$, 7, 8, 9, 9 and 9) were compared. No significant differences were found in pretest scores among the six groups ($F = 0.873$, df = 46). At post-test, two groups had made significant gains, two had non-significantly higher scores and two had non-significantly lower scores[3]
Evaluative (within individual across time)	The same study found that the range of change in individual pre–post-test scores was from 15 to –12[3] The Form has been responsive to participation in an education programme and sensitivity groups (refs 2, 3; also see Construct validity).
Acceptability	
Number of items	89
Administration method	Multiple-choice questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1971, *Journal of Counseling Psychology* The scale is not for rent or sale but may be borrowed for specific research purposes from S. Danish[3]
Web or scanning option	No details
Training details	No details
Administration/process details	Participants view videotape and kinescope excerpts of actual counselling sessions. They are asked to feel whatever emotions the client felt at the end of the excerpt, and to describe the feeling, by checking one of three responses to each statement item
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Binary
Normative data	No details

continued

Notes

The research literature refers to this version as Form B,[2] the Affective Sensitivity Scale[3,7] and Form C[9]

Higher scores have been obtained from using videotapes, as compared to kinescope. This is thought to be due to the better sound quality of the former[1,3]

Other uses of the ASS in research include:

An investigation of the concurrent validity of the Self-disclosure Questionnaire with students in techniques of counselling[8]

An examination of the influence of leader empathy (affective sensitivity), participant motivation to change, and leader–participant relationship changes in affective sensitivity of T-group participants[9]

A study of the meaning of self-awareness in correctional counsellor training with correctional training[10]

A study of relations among components of the empathic process with students[12]

Investigations with undergraduate social workers of the effects of internal empathy and labelling mood on socially demonstrated empathy[13]

Studies of the role of gender in the empathic process with students[22] and school counsellors[23]

Résumé

Strengths	One-week test–retest reliability, and some split-half reliability assessments have indicated adequate reliabilities. The Form has also been responsive to different sample groups, and to participation in programmes designed to increase sensitivity, which supports its construct validity
Weaknesses	Some split-half reliability assessments have offered only partial support. The form has not consistently demonstrated convergent validity. Results so far have been largely unsupportive of the Form's predictive validity. While there is some support for its ability to predict peer ratings, all correlations with outcome so far have not been significant. The Form is also long, with 89 items
Areas for further research	Further assessment of psychometric properties

Primary references

1. Campbell RJ. The development and validation of a multiple-choice scale to measure affective sensitivity (empathy). Unpublished doctoral dissertation, Michigan State University; 1967.
2. Campbell RJ, Kagan N, Krathwohl DR. The development and validation of a scale to measure affective sensitivity (empathy). *J Counsel Psychol* 1971;**18**:407–12.
3. Danish SJ, Kagan N. Measurement of affective sensitivity: toward a valid measure of interpersonal perception. *J Counsel Psychol* 1971;**18**:51–4.
4. Kagan N, Krathwohl D, Goldberg AD, Campbell RJ, Schauble PG, Greenberg BS, *et al*. Studies in human interaction: interpersonal process recall stimulated by videotape. Ann Arbor, MI: Michigan State University Educational Publication Services; 1967.
5. Kagan N, Krathwohl D, Goldberg AD, Campbell RJ, Schauble PG, Greenberg BS, *et al*. Affective Sensitivity Scale – Form (C). Unpublished manuscript, Michigan State University; 1968.
6. Kurtz RR, Grummon DL. Different approaches to the measurement of therapist empathy and their relationship to therapy outcomes. *J Consult Clin Psychol* 1972;**39**:106–15.
7. Kagan N, Schneider J. Toward the measurement of affective sensitivity. *J Counsel Devel* 1987;**65**:459–64.

Secondary references

8. Abendroth WR, Horne AM, Ollendick DG, Passmore JL. Validity of the Self-Disclosure Questionnaire as a measure of counselor effectiveness. *J Counsel Psychol* 1987;**24:**534–76.
9. Danish SJ. Factors influencing changes in empathy following a group experience. *J Counsel Psychol* 1971;**18**:262–67.
10. Grzegorek AE, Kagan N. A study of the meaning of self-awareness in correctional counselor training. *Crim Just Behav* 1974;**1**:99–123.
11. Ham MD. The effects of the relationship between client behavior and counselors' predicted empathic ability upon counselors' in-session empathic performance: an analogue study. Unpublished doctoral dissertation, University of Rochester; 1980.
12. Harman JI. Relations among components of the empathic process. *J Counsel Psychol* 1986;**33**:371–6.
13. Jackson E. Internal empathy, cognitive labeling and demonstrated empathy. *J Hum Educ Dev* 1986;**24**:104–15.
14. Jackson EC, Ahrons C. The relationship of emotional sensitivity to interpersonal skills and practice area speciality. *J Educ Soc Work*, 1984;**21**:74–84.

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

15. Kagan (Klein) H. Interpersonal process recall: influencing human interaction. In Watkins CE Jr, editor. *Handbook of psychotherapy supervision*. New York: John Wiley and Sons; 1997. pp. 296–309.
16. Kagan N, Burke JB, Lieberman M. *The use of physiological recall to develop interpersonal effectiveness in medical students*. Final report. Washington, DC: National Institute of Mental Health; 1982.
17. Kagan NI, Holmes M, Kagan (Klein) H. (Eds.). *Interpersonal process recall manual*. Houston, TX: Mason Media; 1995.
18. Kagan N, Krathwohl DR, Farquhar WW. *Interpersonal process recall: stimulated recall by videotape in exploratory studies of counselling and teacher-learning*. Final report, NDEA Title VII. East Lansing, MI: Michigan State University; 1965.
19. Kirk WG, Thomas AH. A brief inservice training strategy to increase levels of empathy of psychiatric nursing personnel. *J Psychiatr Treat Eval* 1982;**4**:177–9.
20. Kurtz RR., Grummon DL. Different approaches to the measurement of therapist empathy and their relationship to therapy outcomes. *J Consult Clin Psychol* 1972;**39**:106–15.
21. Lieberman MG. Psychophysiological correlates of measures of empathy. Unpublished doctoral dissertation, Michigan State University; 1981.
22. Olesker W, Balter L. Sex and empathy. *J Counsel Psychol* 1972;**19**:559–62.
23. Smith Petro C, Hansen JC. Counselor sex and empathic judgment. *J Counsel Psychol* 1977;**24**:373–6.

A3
Affective Sensitivity Scale (ASS) – Form D

General details	
Authors	Kagan N, Schneider J
Language	English
Country of publication/development	USA
Publication date	1987
Publisher	NA
Purpose and overview	
Form D is a version of the ASS. Developed from IPR, the purpose of the ASS is to test the participant's ability to detect and identify the immediate affective state of another (affective sensitivity). After viewing an excerpt of counselling, participants are asked to feel the affective state of the client, and describe it. For each item, the participant chooses one of three responses (one of which correctly identifies the client's feelings, and two are incorrect) (see Affective Sensitivity Scale – Form A and Form C). Form D was designed as an improved version of Form C[2]	
Theoretical orientation	Interpersonal theory
Population details	Non-clinical adults
Perspective	Self-report
Measure used by	Researchers and the helping professions
Other versions	Forms A, B, C, D-80, E-80, E-A-2, F and H
Notes	Form D participant groups include teachers, undergraduate students and psychiatric nurses[2]
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: empathy/sensitivity	
The therapist–client interaction information is derived from a general description of the measure	
Dimensions	
No details	
Reliability	
Form D has demonstrated adequate internal consistency and partial test–retest reliability[2]	
Split-half	No details
Internal consistency	Cronbach's alpha ($n = 3000$) = 0.74 (probability not reported)
Inter-rater	No details
Test–retest	At an interval of <1 week ($n = 20$), 0.64 (probability not given)
Validity	
Assessments of convergent validity have yielded mixed, but largely unsupportive results. Correlations between Form D and the earlier Form C were extremely low and of zero order. With teacher and undergraduate samples, Form D was found to be unrelated to observer-rated behaviours. Form D demonstrated partial convergent validity when correlated with the Feeling scale of the Myers–Briggs Type Indicator (Myers and Briggs, 1976; see ref. 2)	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Construct	No details
Convergent	Form D correlated 0.43 (p not given) with the Myers–Briggs Type Indicator Feeling Scale[2]
	Correlations between Form D and the earlier Form C were extremely low and of zero order[2]
	No relationship was found between Form D and observer-rated behaviours in teacher and student participant samples[2]
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	Significant differences on Form D have been found between IPR-trained and control group of psychiatric nurses (n = 40) (Kirk and Thomas, 1982; see ref. 2)
	Form C identified high empathisers, as judged by the criteria developed by Truax and Carkuff (1967; see ref. 2)
Evaluative (within individual across time)	No details
Acceptability	
Number of items	No details
Administration method	Multiple-choice questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1987, *Journal of Counseling and Development*
Web or scanning options	No details
Training details	No details
Administration/process details	Participants view videotape and kinescope excerpts of actual counselling sessions. They are asked to feel whatever emotions the client felt at the end of the excerpt, and to describe the feeling, by checking one of three responses to each statement item (see Form A and Form C)
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Binary
Normative data	No details
Notes	
While intended as an improved version of Form C, Form D may be a different instrument. Form D apparently requires more sensitivity to the thoughts and motives, rather than to the feelings of the client[2]	

continued

Résumé

Strengths	Form D has demonstrated adequate internal consistency[2]
Weaknesses	There is only partial support for the Form's test–retest reliability.[2] Correlations so far do not support the convergent validity of Form D: it does not adequately converge with the earlier version (Form C); and correlations with the Briggs Type Indicator Feeling Scale and ref. 2 and observer-rated behaviours were partial and inadequate, respectively[2]
Areas for further research	The authors completely revised Form D, creating a new Form D-80.[2] Therefore, further research with Form D is not necessary

Primary references

1. Kagan NR, Schneider J. Affective Sensitivity Scale, Forms D and E: examiner's manual. Unpublished manuscript, Michigan State University; 1980.
2. Kagan N, Schneider J. Toward the measurement of affective sensitivity. *J Counsel Dev* 1987;**65**:459–64.
3. Werner D. The structure, reliability and validity of the Affective Sensitivity Scale (Form D): a measure of a component of empathy. Unpublished doctoral dissertation, Michigan State University, East Lansing; 1977.

Secondary references

None

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

A4
Affective Sensitivity Scale (ASS) – Form D-80

General details	
Authors	Kagan N, Schneider J
Language	English
Country of publication/development	USA
Publication date	1987
Publisher	NA
Purpose and overview	
Form D-80 is a version of the ASS. Developed from IPR, the purpose of the ASS is to test the participant's ability to detect and identify the immediate affective state of another (affective sensitivity). After viewing an excerpt of counselling, participants are asked to feel the affective state of the client, and describe it (see Affective Sensitivity Scale – Form A and Form C) Form D-80 is a revision of Form D. In addition to rewriting items so that they were more clearly focused on covert affect rather than covert thoughts, the authors developed new subscales (see Dimensions).[1] Form D-80 was developed with Form E-80 and the two have roughly equivalent scenes and multiple-choice questions. The purpose of the parallel forms is to minimise practice effects when used in pretest–post-test experimental design[1]	
Theoretical orientation	Not specified
Population details	Non-clinical adults
Perspective	Self-report
Measure used by	Researchers and for training and supervision in the helping professions
Other versions	ASS Forms A, B, C, D, E, E-80, E-A-2, F and H
Notes	Several groups of training-in-training ($n = 212$) were used in revising Form D to Form D-80
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: empathy/sensitivity The therapist–client interaction information is derived from a general description of the measure	
Dimensions	
Between	28 items. The subscale is an attempt to measure interpersonal sensitivity and consists of all the items that involve the question 'What was [person] × feeling about [person] Y?'[1]
Within	35 items. The subscale is an attempt to measure intrapersonal sensitivity and consists of all the items that involve the question 'What was [person] × feeling at the end of the scene?'[1]
Adult	48 items in which the question refers to the feeling reaction of an adult in the filmed encounter[1]
Child	15 items whose questions refer to the feeling reactions of any child who appeared in the filmed encounter[1]
Male	35 items in which the question refers to a man in the filmed encounter[1]
Female	28 items in which the question refers to a woman in the filmed encounter[1]
Total	The total score includes all items on Form D-80 ($n = 63$)[1]
Emotional accuracy scores	A rating of the primary emotion (anger, anxiety, guilt, distrust, happiness, sadness or helplessness) being expressed in the chosen alternative to each item

continued

Reliability	
Form D-80 has demonstrated partial internal consistency as measured by Cronbach's alpha with a sample of 56[1]	
Split-half	No details
Internal consistency	$r = 0.64$ (probability not reported)[1]
Inter-rater	No details
Test–retest	No details
Validity	
The Form has face validity as the items were written to be more clearly focused on covert affect rather than covert thoughts and the authors developed new subscales (see Dimensions)[1]	
ASS Form D-80 was correlated with its parallel Form E-80 and the resulting coefficients were either partial or inadequate.[1] Among 96 medical students, scores on Form D-80 combined with E-80 converged with tutor-rated empathy or likelihood of seeking help (statistical analyses not reported)	
Face	The items of Form D-80 were written to be more clearly focused on covert affect, rather than covert thoughts[1] Form E-80 was developed alongside a roughly equivalent Form D-80 with the intent that the two be used in pretest–post-test experimental design to minimise practice effects
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	When Forms D-80 and E-80 were correlated, r ranged from 0.28 to 0.67 (p not reported, $n = 210$)[1] High and low scorers on empathy or the likelihood of seeking help (rated by their tutors) were also high and low scorers on the combined Forms of D-80 and E-80 ($n = 96$)[1]
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	63
Administration method	Multiple-choice questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Feasibility	
Copyright	1987, *Journal of Counseling and Development*
Web or scanning options	No details
Training details	No details
Administration/process details	Participants view videotape and kinescope excerpts of actual counselling sessions. They are asked to feel whatever emotions the client felt at the end of the excerpt, and to describe the feeling, by checking one of three responses to each statement item (see Form A and Form C)
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Binary
Normative data	No significant differences have been found between men and women, although each gender tends to score slightly higher in sensitivity to the opposite gender. In a comparison of groups with high and low mean ASS Form D-80 scores, the following significant differences in scale responses were found: High scorers fared better on the scale items between people High scorers showed more sensitivity to adults; low scorers did better on sensitivity to children High scorers showed slightly more sensitivity to men; low scorers did better on sensitivity to women
Normative data	Low scorers were more likely to identify happiness, sadness and helplessness when not present, and to miss helplessness and sadness when they were present[1]
Notes	
The ASS is used for supervision purposes in helping professions	
Résumé	
Strengths	Compared to earlier versions of the ASS, Form D-80 is more focused on affect than thoughts[1]
Weaknesses	Assessment of the Form suggests that internal consistency is only partial.[1] Forms D-80 and E-80 were designed to be parallel, yet their relationship is unclear, with correlation coefficients ranging from inadequate to adequate.[1] While shorter than earlier versions, Form D-80 is still long, with 63 items.[1]
Areas for further research	Further assessments of the Form's psychometric properties, particularly its relationship with Form E-80
Primary reference	
1. Kagan N, Schneider J. Toward the measurement of affective sensitivity. *J Counsel Dev* 1987;**65**:459–64.	
Secondary references	
None	

A5
Affective Sensitivity Scale (ASS) – Forms E-80 and E-A-2

General details	
Authors	Kagan N, Schneider J
Language	English
Country of publication/development	USA
Publication date	1987
Publisher	NA
Purpose and overview	
Form E-80 is a version of the ASS. Developed from IPR, the purpose of the ASS is to test the participant's ability to detect and identify the immediate affective state of another (affective sensitivity). After viewing an excerpt of counselling, participants are asked to feel the affective state of the client, and describe it (see Affective Sensitivity Scale – Form A and Form C)	
Form E-80 is a revision of Form E. In addition to rewriting items so that they were more clearly focused on covert affect rather than covert thoughts, the authors developed new subscales (see Dimensions)[1]	
Form E-80 was developed with Form D-80 and the two have roughly equivalent scenes and multiple-choice questions. The purpose of the parallel forms is to minimise practice effects when used in pretest–post-test experimental design[1]	
Form E-A-2 preceded Form D-80[1]	
Theoretical orientation	Interpersonal theory
Population details	Non-clinical adults
Perspective	Self-report
Measure used by	Researchers and for training and supervision in the helping professions
Other versions	ASS Forms A, B, C, D, D-80, E, F and H
Notes	Several groups of training-in-training ($n = 2461$) were used in developing Form E-80
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: empathy/sensitivity	
The therapist–client interaction information is derived from a general description of the measure	
Dimensions	
Between	24 items. The subscale is an attempt to measure interpersonal sensitivity and consists of all the items that involve the question 'What was [person] × feeling about [person] Y?'[1]
Within	34 items. The subscale is an attempt to measure intrapersonal sensitivity and consists of all the items that involve the question 'What was [person] × feeling at the end of the scene?'[1]
Adult	41 items in which the question refers to the feeling reaction of an adult in the filmed encounter[1]
Child	16 items whose questions refer to the feeling reactions of any child who appeared in the filmed encounter[1]
Male	26 items in which the question refers to a man in the filmed encounter[1]
Female	31 items in which the question refers to a woman in the filmed encounter[1]
Total	The total score includes all items on Form E-80 ($n = 57$)[1]
Emotional accuracy scores	A rating of the primary emotion (anger, anxiety, guilt, distrust, happiness, sadness or helplessness) being expressed in the chosen alternative to each item

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
ASS Form E-80 has demonstrated partial test–retest reliability[5]	
ASS Form E-A-2 has shown partial internal consistency as measured by Cronbach's alpha[1]	
Test–retest reliabilities have been partial and adequate for Form E-A-2 as a whole, and partial for the Within subscale[8]	
Split-half	No details
Internal consistency	E-A-2 Cronbach's alpha $r = 0.58$ ($n = 1200$)[1]
Inter-rater	No details
Test–retest	E-80 7-week test–retest ($n = 44$), $r = 0.62$[5]
	E-A-2 test–retest $r = 0.69$ ($n = 23$), 0.77 ($n = 20$) and 0.52 ($n = 25$)[1]
	E-A-2 Within subscale test–retest 0.60 ($p < 0.01$)[8]
Validity	
In assessments of ASS Form E-80's convergent validity, it was correlated with ASS Forms C and D-80. The coefficients with Form C were adequate, and coefficients with Form D-80 ranged from inadequate to adequate[1]	
Form E-80's convergent validity has been inadequate with Carkhuff (1969) empathy ratings[11] and the Hogan Empathy Scale (1969)[4]	
Only two aspects of Form E-80 showed a significant relationship with the Barrett-Lennard Relationship Inventory (BLRI, 1978).[11] Correctly identifying guilt, incorrectly correctly identifying anger, both negatively correlated with the BLRI. The meaning of the correctly identifying guilt correlation is unclear, while incorrectly identifying anger correlated in the expected direction[11]	
A significant relationship was found between Form E-A-2 and ego development, which indicates support for convergent validity[2]	
ASS Form E-A-2 showed partial convergent validity when correlated with peer ratings of warmth, understanding and openness.[6] The Within subscale partially converged with behavioural interview ratings[8]	
Face	Form E-80 was developed alongside a roughly equivalent Form D-80 with the intent that the two be used in pretest–post-test experimental design to minimise practice effects
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	ASS Form E-80 and Form C were correlated and $r = 0.56$ (p not reported)[1]
	When Forms D-80 and E-80 were correlated, r ranged from 0.28 to 0.67 (p not reported, $n = 210$)[1]
	No relationship was found between Form E-80 and Carkhuff (1969) ratings of empathy[11]
	High scores on Form E-80 (correctly identifying guilt) correlated with the BLRI –0.34 ($p < 0.01$, $n = 90$)[11]
	High scores on Form E-80 (incorrectly identifying anger) correlated with the BLRI –0.36, ($p < 0.01$)[11]
	From E-A-2 converged with peer ratings of warmth, understanding and openness ($r = 0.45$, $p < .001$, $n = 44$)[6]
	Form E-A-2 'within' subscale converged with behavioural interview ratings ($r = 0.40$, $p < 0.05$)[8]
	A significant relationship was found between ego development and Form E-A-2 scores ($p < 0.05$)[2]
Discriminant	No details
Factor structure	No details

continued

Responsiveness	
Discriminative (between individuals)	Significant differences in E-80 scores were found between participants who were ranked by peers in terms of someone who would make a good counsellor and be helpful with an emotional or interpersonal problem ($p = 0.0112$, 0.0364)[7]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	E-80 has 57 items and E-A-2 has 65
Administration method	Multiple-choice questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1987, *Journal of Counseling and Development*
Web or scanning options	No details
Training details	No details
Administration/process details	Participants view videotape and kinescope excerpts of actual counselling sessions. They are asked to feel whatever emotions the client felt at the end of the excerpt, and to describe the feeling, by checking one of three responses to each statement item
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Binary
Normative data	No significant differences have been found between men and women, although each gender tends to score slightly higher in sensitivity to the opposite gender. In a comparison of groups with high and low mean ASS Form E-80 scores, the following significant differences in scale responses were found: High scorers fared better on the scale items between people High scorers showed more sensitivity to adults; low scorers did better on sensitivity to children High scorers showed slightly more sensitivity to men; low scorers did better on sensitivity to women
Normative data	Low scorers were more likely to identify happiness, sadness and helplessness when not present, and to miss helplessness and sadness when they were present[1]
Notes	
The ASS is used for supervision purposes in helping professions Uses of the ASS Form E-80 in research include: A study of the relationship between empathy and ego development[2] A multiple regression approach to the study of the relationship between peer-rated therapeutic talent and affective sensitivity with undergraduates[6] A study of the effects of the relationship between client behaviour and training 'predicted empathic ability upon training' in-session empathic performance[4] A study of the psychophysiological correlates of measures of empathy[11] A study of the role of defence mechanisms, relaxation and guided imagery in affective sensitivity[12]	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Résumé	
Strengths	*Form E-80* Compared to earlier versions of the ASS, Form E-80 is more focused on affect than thoughts, and is shorter. Large sample sizes were employed in the development of the form. Form E-80 has shown responsiveness to peer rankings of counselling qualities and helpfulness with emotional or interpersonal problems[7] *Form E-A-2* Scores on the Form adequately converged with ego development[2]
Weaknesses	*Form E-80* Form E-80 internal consistency[1] and test–retest reliability coefficients are only partial.[5] Form E-80 failed to demonstrate adequate convergent validity when correlated with several measures: the BLRI (1978) and Carkhuff (1969) empathy ratings;[11] the Hogan Empathy Scale (1969);[4] ASS Forms C and D-80.[1] Forms D-80 and E-80 were designed to be parallel, yet their relationship is unclear, with correlation coefficients ranging from inadequate to adequate.[1] While shorter than earlier versions, Form E-80 is still long, with 57 items[1] *Form E-A-2* The form is long, with 65 items. It has only partially correlated with peer ratings of warmth, understanding and openness,[6] and behavioural interview ratings[8]
Areas for further research	Further examination of psychometric properties, particularly the relationship between Forms E-80 and D-80 and convergence with other measures of empathy

Primary reference

1. Kagan N, Schneider J. Toward the measurement of affective sensitivity. *J Counsel Dev* 1987;**65**:459–64.

Secondary references

2. Carlozzi AF, Gaa JP, Liberman DB. Empathy and ego development. *J Counsel Psychol* 1983;**30**:113–16.
3. Ginsburg MF. A prescriptive counselor training approach for enhancing accurate awareness of client affect. Unpublished doctoral dissertation, Syracus University; 1981.
4. Ham MD. The effects of the relationship between client behavior and counselors' predicted empathic ability upon counselors' in-session empathic performance: an analogue study. Unpublished doctoral dissertation, University of Rochester; 1980.
5. Jackson EC. Sensitivity to emotions and behavioral empathy in the initial counseling interview. *J Hum Educ Dev* 1984;**22**:107–14.
6. Jackson E. The relationship of peer rated therapeutic talent and affective sensitivity: a multiple regression approach. *Psychology Q J Hum Behav* 1985;**22**:29–35.
7. Jackson E. Affective sensitivity and peer selection of counseling potential. *Counsel Educ Supervis* 1986;**25**:230–6.
8. Jackson E. Internal empathy, cognitive labeling and demonstrated empathy. *J Hum Educ Dev* 1986;**24**:104–15.
9. Jackson EC, Ahrons C. The relationship of emotional sensitivity to interpersonal skills and practice area speciality. *J Educ Soc Work* 1984;**21**:74–84.
10. Kagan N, Burke JB, Lieberman M. The use of physiological recall to develop interpersonal effectiveness in medical students. Final report. Washington, DC: National Institute of Mental Health; 1982.
11. Lieberman MG. Psychophysiological correlates of measures of empathy. Unpublished doctoral dissertation, Michigan State University; 1981.
12. Milhouse J. The role of defense mechanisms, relaxation and guided imagery in affective sensitivity. Unpublished doctoral dissertation, Michigan State University; 1982.

A6
Kagan Affective Sensitivity Scale (KASS) – Form H

General details

Author	Kagan N
Language	English
Country of publication/development	USA
Publication date	1994
Publisher	NA

Purpose and overview

Form H is the current version of the Kagan Affective Sensitivity Scale (KASS),[2] although the previous versions are referred to under Affective Sensitivity Scale (ASS)

Developed from IPR, the purpose of the ASS is to test the participant's ability to detect and identify the immediate affective state of another (affective sensitivity) (see Affective Sensitivity Scale – Form A and Form C)

Theoretical orientation	Interpersonal theory
Population details	Non-clinical adults
Perspective	Self-report
Measure used by	Researchers and for training and counselling in the helping professions
Other versions	ASS Forms A, B, C, D, E, E-80, E-A-2 and F

Areas of therapist–patient interaction addressed: Map

Therapist engagement: empathy/sensitivity

The therapist–client information is derived from a general description of the measure

Dimensions

Empathy subscale[2]

Reliability

The internal consistency of KASS Form H, as measured by Kuder–Richardson correlation coefficient, is adequate[1]

Split half	No details
Internal consistency	Kuder–Richardson correlation coefficient = 0.78 for the total scale[1]
Inter-rater	No details
Test–retest	No details

Validity

KASS Form H has predicted poor performance in the professions of counselling, teaching, medicine and law[1]

Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	Low scores on KASS Form H appear to predict poor performance in counselling, teaching, medical and legal professions, although high scores do not seem to predict future success in these fields[1]
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Responsiveness	
Discriminative (between individuals)	The empathy subscale of the KASS Form H discriminated against two groups, one using only recall and one using the IPR manual, with the former scoring significantly higher[4]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	No details
Administration method	Multiple-choice questionnaire
Time taken-to-complete	Approximately 80 minutes
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1994, Mason Media
Web or scanning options	Machine scored
Training details	No details
Administration/process details	Participants view videotape and kinescope excerpts of actual counselling sessions. They are asked to feel whatever emotions the client felt at the end of the excerpt, and to describe the feeling, by checking one of three responses to each statement item (see Form A and Form C)
Support from measure developers	No details
FAQ facility	For enquiries: e-mail: drhenya@gilisplace.com Fax: +17136690567 (US number)[3]
Precision	
Scale type	Binary
Normative data	No details
Notes	
1003 – KASS Package costs $395 with an additional fee of $30 for scoring and shipping and is available from: Mason Media, Inc., Dr Henya Shaun-Klein, PO Box 20712, Houston, TX 77225-0712, USA[3]	
Résumé	
Strengths	The Form has demonstrated adequate test–retest reliability[1]
Weaknesses	The form is expensive and, with a completion time of 80 minutes, fairly lengthy
Areas for further research	Examination of psychometric properties: few areas have been addressed

Primary reference

1. Kagan NI. *Kagan Affective Sensitivity Scale, Form H*. Examiner's Manual. Houston, TX: Mason Media; 1994.

Secondary references

2. Kagan (Klein) H. Interpersonal process recall: influencing human interaction. In Watkins CE Jr, editor. *Handbook of psychotherapy supervision*. John Wiley and Sons; 1997. pp. 296–309.
3. URL: http://www.kaganklein.com/IPR-Kass/Kass
4. Spaulding J. Increasing empathy and interpersonal skills in community college students through the use of interpersonal process recall. Unpublished candidacy paper, University of Houston; 1993.

A7
Agnew Relationship Measure (ARM)

General details	
Author	Agnew-Davies R
Language	English
Country of publication/development	UK
Publication date	1998
Publisher	NA
Purpose and overview	
The ARM is a UK self-report questionnaire designed to describe components of the alliance in a language designed to be acceptable across a wide range of theoretical orientations	
Theoretical orientation	Pan-theoretical
Population details	Adults
Perspective	Therapist and client
Measure used by	Therapists/counsellors
Other versions	12-item version
Areas of therapist–patient interaction addressed: Map	
Therapy context: influence; responsibilities	
Roles: friend/companion; expert/authority/leader	
Therapist engagement: empathy/sensitivity; warmth; genuineness; respect; support/tolerance; openness; listening; hope/encouragement	
Patient engagement: motivation; commitment; intentions; attraction	
Framework: convergent; complementary; reciprocal; collaborative/participative/involving; congruent; controlling; structuring; directive; challenging; flexible/rigid	
Therapeutic techniques: responsiveness/receptiveness/attunement; exploration	
Threats to the relationship: defensive; critical; fear; resistance; withdrawal	
Outcomes: general satisfaction; working alliance; cohesion; emotional expression	
Information derived from items	
Dimensions	
Bond	Bond has six items and concerns the friendliness, acceptance, understanding and support in the relationship
Partnership	Partnership has four items and concerns working jointly on therapeutic tasks
Confidence	Confidence has seven items and concerns optimism and respect for the therapist's professional competence
Openness	Openness has five items and concerns the client's freedom to disclose personal concerns without fear or embarrassment
Client initiative	Client initiative has four items and concerns the client's taking responsibility for the direction of the therapy

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
Using Cronbach's alpha as a reliability estimate, the four scales of bond, openness, confidence and partnership demonstrated adequate internal consistency. The client initiative scale demonstrated partial adequacy, but in one study[2] the therapist's scale was demonstrated to be inadequate	
Split-half	No details
Internal consistency	Alpha ranged from 0.77 to 0.87[1] and from 0.73 to 0.89[2] The internal consistency for the client initiative is lower than for the other scales: α 0.55 for both client and therapist[1] and 0.41 to 0.59[2]
Inter-rater	NA
Test–retest	No details
Validity	
The ARM demonstrates adequate convergent validity with the Working Alliance Inventory (WAI) within perspective Evidence for the predictive validity of the ARM is mixed. There is a lack of independence between the scales, suggesting a degree of overlap between the dimensions	
Face	The ARM avoids content that describes technique or presumes links to outcome and uses a language suitable for most therapeutic approaches[1]
Content	The ARM samples broad content areas of alliance defined in previous work, including internally consistent scales for openness and confidence in addition to the more standard content for bond and partnership[1]
Criterion (a) concurrent	No details
Criterion (b) predictive	Some aspects of the alliance as measured by the ARM were related to client outcome at the end of treatment.[4] The strength of this association varied according to assessment measures, ARM scales and the session number when the alliance was measured. (See ref. 4 for further details)
Convergent	The convergent validity of the ARM with the WAI was demonstrated in ref. 2. This study used data from two studies to assess convergent validity: the Collaborative Psychotherapy Project (CPP) and the Second Sheffield Psychotherapy Project (SPP2). Convergent validity was assessed at the dyad level and session level (see ref. 2 for details). The correlations of the ARM bond, partnership, and confidence scales with the WAI bond, goals and tasks within client therapist perspectives in CPP were all in the 0.80s and 0.90s. The dyad-level correlations were weaker in the SPP2 sample. Within-perspective session-level correlations for the core alliance scales in CPP were also strong, but lower (0.54–0.70 for clients; 0.57–0.85 for therapists). Between-perspective convergent validity was moderate (see Tables 2 and 3 in ref. 2 for full details)
Discriminant	No details
Factor structure	Simultaneous components analysis (SCA) was used to extract six components for each perspective (client and therapist). All the items in both the bond and partnership scales loaded highly on component 1. In the confidence scale, all items loaded highly on component 1 for the clients' ratings, but were separate (component 3) in the therapists' ratings. The openness scale's items comprised a distinct component in each perspective (client component 3; therapist component 2). On the client initiative scale, items 4 and 25 comprised a distinct two-item component, and internal consistency would be better served by including these two items only. However, the authors preferred the broader scope afforded by including items 23 and 11. Items 18 and 28 were excluded because they were used differently by clients and therapists Intercorrelations between the bond, partnership and confidence scales were fairly high within therapist and client perspectives, indicating a degree of overlap between these scales

continued

Responsiveness	
Discriminative (between individuals)	The ARM distinguishes between clients on age.[1] Alliances were rated as slightly stronger by older clients on the bond, partnership, confidence and openness scales. Therapists rated older clients as relatively higher on openness. Therapists rated women as slightly higher than men in openness. Although the alliance was found to be positive in both cognitive behavioural (CB) and psychodynamic interpersonal (PI) treatments (means on all five scales were above the midpoint) there was a trend for the ARM to distinguish clients on the basis of treatment condition.[1,3] There were two nominally significant differences between the treatments: clients reported a stronger partnership in CB (mean = 6.06) than in PI (mean = 5.56, $F_{1,75} = 9.27$, $p = 0.003$) and a slightly greater confidence in CB (mean = 5.99) than in PI (mean = 5.60, $F_{1,75} = 6.25$, $p = 0.015$)
Evaluative (within individual across time)	The ARM has been used to measure change over time[2,4]
Acceptability	
Number of items	28
Administration method	Self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1998, British Psychological Society
Web or scanning options	No details
Training details	No details
Administration/process details	The ARM is to be completed after each therapy session. Instructions on the form read 'Thinking about today's meeting, please indicate how strongly you agree or disagree with each statement'
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Seven-point Likert scale anchored from 1 (strongly disagree) to 7 (strongly agree). Higher scoring indicates better alliance
Normative data	No details
Notes	
Two pilot versions preceded the 28-item ARM (see Agnew R. Conceptual and methodological issues in the development of a measure of psychotherapeutic relationships. PhD thesis, University of Sheffield; 1996)	
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Résumé	
Strengths	The ARM incorporates the main content areas identified in previous work, including internally consistent scales for openness and confidence, in addition to the more standard content for bond and partnership. The ARM avoids content that describes technique or presumes links to outcome. It offers a relatively simple format in language appropriate for most therapeutic approaches. The psychometric properties of the ARM have been adequately demonstrated in a number of studies
Weaknesses	There are two main drawbacks: (1) There is a lack of statistical independence between the conceptually distinct bond and partnership scales; this is a drawback common to other alliance instruments. (2) The demographic and diagnostic range of the clients is restricted; to date, the measure has only been used with professional, managerial and other white-collar workers referred for treatment of depression
	At 28 items the measure is overly long for use in routine practice
Areas for further research	Application of the ARM to other sorts of data and settings, which will enhance the measure's validity and usefulness
	Validation work on the ARM short version (12 items)

Primary references

1. Agnew-Davies R, Stiles WB, Hardy GE, Barkham M, Shapiro DA. Alliance structure assessed by the Agnew Relationship Measure (ARM). *B J Clini Psychol* 1998;**37**:155–72.
2. Stiles WB, Agnew-Davies R, Barkham M, Culverwell A, Goldfried MR, Halstead J, *et al*. Convergent validity of the Agnew Relationship Measure and the Working Alliance Inventory. *Psychol Assess* 2002;**14**:209–20.

Secondary references

3. Hardy GE, Agnew-Davies R, Stiles WB, Barkham M. When and why does cognitive behavioural treatment appear more effective than psychodynamic-interpersonal treatment? Discussion of findings from the Second Sheffield Psychotherapy Project. *J Ment Health* 1998;**7**:179–90.
4. Stiles WB, Agnew-Davies R, Hardy GE, Barkham M, Shapiro DA. Relations of the alliance with psychotherapy outcome: findings in the second Sheffield Psychotherapy Project. *J Consult Clin Psychol* 1998;**66**:791–802.

B1
Barrett-Lennard Relationship Inventory (BLRI or RI)

General details

Author	Barrett-Lennard GT
Language	English and several other languages
Country of publication/development	USA and Australia
Publication date	1962
Publisher	NA (see copyright details)

Purpose and overview

To measure four dimensions of the interpersonal relationship adapted from Rogers' (1957, 1959) conception of the necessary conditions for therapeutic personality change. The RI has been used to test the predictive ability of the therapeutic relationship on outcome. The four dimensions are therapist empathy, regard, unconditionality of regard and congruence

Theoretical orientation	Developed from a person-centred/Rogerian perspective, but is for general use in therapy/counselling relationships
Population details	Clinical adults (attending individual and group therapy), clinical adolescents, non-clinical adults, non-clinical children, non-clinical students (undergraduates and graduates), parents. Ref. 1 details the wide range of populations involved in BLRI studies. Service settings include secondary care (e.g. refs 3, 5) and educational settings (e.g. refs 8, 10, 11, 14, 15). Information about the BLRI for use in family/systemic therapy is also available[12]
Perspective	Patient self-report, therapist rated and independent observer/rater
Measure used by	Psychotherapists (e.g. refs 3, 5); counsellors (e.g. refs 8, 10, 11); psychologists (e.g. ref. 14); trainees/students (e.g. ref. 15); nurses (e.g. ref. 7)[1]
Other versions	The main versions are for clients in individual counselling/therapy (OS-64) and a corresponding form for therapists with items worded in the first person (MO-64). In addition there are three alternative forms: Group therapy: Form OS-G-64 is for use in assessing an individual's perception of the relationship conditions presented by a group as a whole (Barrett-Lennard, 1972) Family, friends, colleagues: experimental form OS-S-42 Teacher–pupil: Scheuer (1971)[17]
Notes	It can be and has been used with clinical and non-clinical samples, covering a range of interpersonal relationships

Areas of therapist–patient interaction addressed: Map

Therapist engagement: empathy/sensitivity; respect; openness; support/tolerance; warmth

Framework: (congruent)

Therapeutic techniques: responsiveness/receptiveness/attunement

Non-verbal communication

Outcomes: general satisfaction (satisfaction); achieving a working relationship; expression of feelings

Information derived from example items

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Dimensions	
Empathy Regard	In each dimension, eight items are expressed positively and eight negatively. The 64 items also give an overall total score
Unconditionality of regard Congruence	The four scales of the RI have been shown to be related, but separate aspects of the relationship and were not developed to be used independently. A number of studies have, however, used the empathy scale alone, to compare it with other empathy scales and to test its predictive validity
Notes	A fifth variable, 'willingness to be known', was included in the original BLRI (Barrett-Lennard, 1962), but dropped in a revision of the measure (Barrett-Lennard, 1969), because of its lack of predictive power in relation to therapeutic outcome
Reliability	
The mass of literature on the BLRI on the whole indicates a high degree of reliability of the measure, including the internal consistency of the measure and its stability over time	
Split-half	Eleven studies using the split-half method with the four subscales are reported.[1] The reliabilities across the studies, for each subscale and total scores, are uniformly high. Mean reliability coefficients are: empathy 0.84; regard 0.91; unconditionality of regard 0.74; congruence 0.88; total 0.91 Other studies supporting the internal consistency of the RI: refs 5–7
Internal consistency	Three studies used the alpha coefficient method to test internal consistency of the four subscales, two of which tested the Inventory total score.[1] The coefficients were: empathy 0.88, 0.88, 0.64; regard 0.91, 0.92, 0.83; unconditionality of regard 0.76, 0.67, 0.73; congruence 0.92, 0.90, 0.80; total 0.95, 0.93. The reported studies include individual and group therapy, and are from actual therapy and therapy analogues. Full details in ref. 1. Other studies supporting the internal consistency of the RI: refs 3, 4, 6 and 7
Inter-rater	Ref. 6 details poor inter-rater reliability scores on the observer forms
Test–retest	Ten studies of test–retest reliability are reported.[1] All show good stability, with mean correlations of: empathy 0.83; regard 0.83; unconditionality of regard 0.80; congruence 0.85; total 0.90. The majority of these studies used relatively short intervals between tests (approx. 1 month), but three studies found a high degree of stability over several months. Studies used client and therapist versions of the RI. Full details of studies in ref. 1. Other studies supporting the test–retest of the RI: refs 3, 5, 6 and 7
Validity	
The mass of literature on this measure on the whole supports the validity of this measure, including its face validity and factor structure. Less evidence for convergent validity was found	
Face	Initial items derived from Rogers' (1957) paper on the conditions of therapy and from Brown's (1954) Relationship Sort.[1,3] Revisions of the items followed consultation with staff at the University of Chicago Counseling Center
Content	A formal content validation procedure was carried out to eliminate non-differential items.[1] Five judges (client-centred therapists) categorised items as positive, negative or neutral for irrelevant or unclear items. There was perfect agreement on all but four items
Criterion (a) concurrent	No details

continued

Criterion (b) predictive	The Relationship Inventory, administered after five therapy interviews, predicted both therapist-rated and client-rated outcome.[5] Correlations between outcome and all client-perceived relationship scales were significant to at least the 5% level. Generally, where both therapist and client inventory scores were high, the majority of clients had better outcomes; where client scores were high but therapist scores low, fewer showed better adjustment; where therapist scores were high, but client scores low, less than half the clients improved substantially; and only a minority improved where both therapist and client scores were low Three studies that have compared the predictive value of the RI with judge-rated relationship measures have found that empathy and regard are correlated with outcome, and that the RI scores are a better predictor than judge-rated measures[1] Other references generally supporting the predictive validity of the BLRI include: 14 and 16, although ref. 1 is more critical
Construct	The Inventory's empathy scales were assessed for construct validity by calculating the coefficients with four other scales that purport to measure empathy.[14] With one exception, the correlations between the BLRI scale and the other measures of empathy were low and insignificant
Convergent	Ref. 7 provides evidence of convergent validity for the BLRI empathy scale with the ECRS: significantly correlated ($r = 0.78$, $p < 0.001$). However, in ref. 14 there were no significant correlations between Inventory's scales and another process measure (Self Exploration in Interpersonal Process Scale; Carkhuff and Berenson, 1967); and in ref. 15 there were no significant relationships between the Inventory scales and the Truax scales
Discriminant	No details
Factor structure	Nine studies have computed the intercorrelations of the client Inventory scales.[1] Empathy, regard and congruence are relatively dependent, while unconditionality of regard is quite independent. As the reliability of the scales is high, the implication is that they are consistently measuring overlapping but separate dimensions of the client-perceived relationship. Five studies using principal components factor analysis, without rotation, found either a one-factor solution with E, C and R loading highly and U with a low loading; or a two-factor solution with E, R and C contributing the most to the first, and U contributing most to the second factor. These studies are criticised for conducting analyses based on interscale correlations without beginning with item intercorrelations, and for not using rotation. Three factor analytic studies that began with item intercorrelations indicate the RI is tapping dimensions that are consistent with Barrett-Lennard's original work on the Inventory. Two studies, which are not directly comparable, offer conflicting evidence as to whether the dimensional structure of the RI varies between populations with varying degrees of psychological disturbance. Further research is needed to clarify this point. Full details of studies in ref. 1, summarised in ref. 2 References detailing, and on the whole supporting the factor structure include 4, 5, 8–12. References questioning the factor structure include 1, 3 and 5
Responsiveness	
Discriminative (between individuals)	Review:[1] Barrett-Lennard (1962) has reported differences in the client-perceived relationship between clients expert vs non-expert therapists (see ref. 5, below). Studies reviewed in ref. 1, however, suggest that the RI is tapping into therapist factors other than expertise, such as maturity, and question Barrett-Lennard's categorisation of 'expert' and 'non-expert' therapists. With the exception of the willingness to be known scale, the Inventory distinguished between expert and non-expert therapists[5] The RI has tapped unidentified differences between male and female observer raters.[6] Snelbecker (1961, 1967) established significant gender effects, with women yielding higher scores
Evaluative (within individual across time)	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Acceptability	
Number of items	64
Administration method	Questionnaire
Time taken to complete	<15 minutes[2]
Flesch reading age	No details
Translations	The BLRI has been translated into several languages (no details)
Access by ethnic minorities	No details
Feasibility	
Copyright	GT Barrett-Lennard. The BLRI and permission for its use can be obtained from GT Barrett-Lennard, The Centre for Studies in Human Relations, 6 Dover Crescent, Wembley Downs, W.A. 6019, Australia
Web or scanning options	No details
Training details	No details
Administration/process details	Questionnaire
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert. Each item is scored 'definitely true' to 'definitely not true' on a six-point Likert-type scale;[1] e.g. 'I feel it is probably true (or not true)', 'I feel it is true (or not true)', 'I strongly feel that it is true (or not true)'. The items are arranged so that every fourth item taps the same variable. Positive and negative items for each dimension are equally distributed between the two halves of the test. Scoring details are provided in ref. 1
Normative data	No details of any established 'norms', but there are substantial published data on this measure
Notes	
The RI originally had 72 items, which had been subject to content validation, split-half and test–retest reliabilities, but not empirical item analysis. The RI was revised to enhance the quality of the items, to make it shorter and easier, and to be adaptable for use with the teacher/pupil relationship. The resulting RI is a 64-item version which has become the regular RI used in research (Barrett-Lennard, 1964) The BRLI has been used in a variety of medical education and clinical situations:[2] the paper reports findings from applications of the BLRI in medicine, including use in nurse–patient relationship studies and medical students' learning This summary[13] refers mainly to the client and therapist versions of the BLRI. Ref. 13 reviews the development, revisions and empirical support for the BLRI, which is largely covered elsewhere. The chapter adds to the existing review literature by providing information on BLRI form OS-S-42, which yields a profile of the respondent's interpersonal world, covering family, 'friend' and work relationships.	
Résumé	
Strengths	In a review of research and concepts surrounding the measurement of facilitative conditions, it is concluded that the BLRI "continues to be the most effective method of measuring the facilitative conditions in a manner that is true to Rogers' theory" (Gelso and Fretz, 1992; p.143, cited in ref. 2) The BLRI takes <15 minutes to complete and has four independently modifiable subscales that can be used independently.[2] The BLRI and its subscales have documented reliabilities that tend to range above 0.80. The BLRI is grounded on an established theoretical foundation, and it allows measurement of an important aspect of care – the quality of the patient–therapist relationship

continued

	The BLRI can measure the relationship from patient, therapist and observer perspectives and is versatile in that it can be applied to many settings, including group, couples and family work
	The BLRI's main strength is its extensive use in field and clinical settings.[16] It has been validated in actual counselling situations more frequently than in analogue situations
Weaknesses	Although there is a mass of literature on the BLRI to support its use, the majority of the literature is now relatively dated.[16] The BLRI's many forms make it difficult to evaluate its overall empirical strength. Also, systematic follow-up of a particular BLRI form has been hampered by researchers' repeated modifications (e.g. Mills and Zytowski, 1967; Claiborn *et al*., 1983)
Areas for further research	Further research is needed to determine whether the dimensional structure of the RI varies between populations with differing degrees of psychological disturbance.[1] It would also be useful if researchers developed a consensus for scale content and then systematically assessed its construct validity[16]

Primary references

1. Gurman AS. The patient's perception of the therapeutic relationship. In Gurman AS, Razin AM, editors. *Effective psychotherapy: a handbook of research.* Oxford: Pergamon Press; 1977. pp. 503–43. (NB. Key review article.)
2. Simmons J, Roberge L, Kendrick SB, Richards B. The interpersonal relationship in clinical practice: the Barrett-Lennard Relationship Inventory as an assessment instrument. *Eval Health Prof* 1995;**18**:103–12.
3. Mills DH, Zytowski DG. Helping relationship: a structural analysis. *J Counsel Psychol* 1967;**14**:193–7.
4. Wiebe B, Pearce WB. An item-analysis and revision of the Barrett-Lennard Relationship Inventory. *J Clin Psychol* 1973;**29**:495–7.
5. Barrett-Lennard GT. Dimensions of therapist response as causal factors in therapeutic change. *Psychol Monogr* 1962;**76**: Whole No. 562, 1–36.
6. Barrett-Lennard GT. The Relationship Inventory: later development and adaptations. *Catalog of Selected Documents in Psychology, 8 MS*. 1978;**1732**:68.
7. Layton JM, Wykle MH. A validity study of four empathy instruments. *Res Nurs Health* 1990;**13**:319–25.
8. Tosi DJ, Frumkin RM, Wilson ME Jr. Intercorrelations of four relationship components of the Barrett-Lennard Relationship Inventory. *Psychol Rep* 1968;**23**:641–42.
9. Walker BS, Little DF. Factor analysis of the Barrett-Lennard Relationship Inventory. *J Counsel Psychol* 1969;**16**:516–21.
10. Lanning, WL, Lemons SL. Another look at the factor structure of the Barrett-Lennard Relationship Inventory. *Measure Eval Guid* 1974;**6**:228–31.
11. Claiborn CD, Crawford JB, Hackman HW. Effects of intervention discrepancy in counseling for negative emotions. *J Counsel Psychol* 1983;**30**:164–71.
12. Ganley RM. The Barrett-Lennard Relationship Inventory (BLRI): current and potential uses with family systems. *Family Proc* 1989;**28**:107–15.
13. Barrett-Lennard GT. The Relationship Inventory now: issues and advances in theory, method and use. In Greenberg SL, Pinsof WM, editors. *The psychotherapeutic process: a research handbook*. New York: Guilford; 1986. pp. 439–76.
14. Kurtz RR, Grummon DL. Different approaches to the measurement of therapist empathy and their relationship to therapy outcomes. *J Consult Clin Psychol* 1972;**39**:106–15.
15. McWhirter JJ. Two measures of the facilitative conditions: a correlation study. *J Counsel Psychol* 1973;**20**:317–20.
16. Ponterotto JG, Furlong MJ. Evaluating counselor effectiveness: a critical review of rating scale instruments. *J Counsel Psychol* 1985;**32**:597–616.

Secondary references

17. Scheuer AL (1971). The relationship between personal attributes and effectiveness in teachers of the emotionally disturbed. *Exceptional Children* 1971;**37**:723–31.
18. Arachtingi BM, Lichtenberg JW. Self-concept and self-esteem as moderators of client transference. *Psychotherapy* 1999;**36**:369–79.
19. Borrego RL, Chavez EL, Titley RW. Effect of counselor technique on Mexican-American and Anglo-American self-disclosure and counselor perception. *J Counsel Psychol* 1982;**29**:538–41.
20. Bradley FO. A modified interpersonal process recall technique as a training model. *Counsel Edu Supervis* 1974;**14**:34–9.
21. Clark JV, Culbert SA. Mutually therapeutic perception and self-awareness in a T group. *J Appl Behavioral Sci* 1965;**1**:180–94.
22. Curtis JM. The effect of therapist self-disclosure on patients' perceptions of empathy, competence and trust in an analogue psychotherapeutic interaction. *Psychotherapy* 1982;**19**:54–62.

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

23. Ellickson JL. Representational systems and eye movements in an interview. *J Counsel Psychol* 1983;**30**:339–45.
24. Esse JT, Wilkins W. Empathy and imagery in avoidance behavior reduction. *J Consult Clin Psychol* 1978;**46**:202–3.
25. Feldstein JC. Effects of counselor sex and sex role and client sex on clients' perceptions and self-disclosure in a counseling analogue study. *J Counsel Psychol* 1979;**26**:437–43.
26. Feldstein JC. Counselor and client sex pairing: the effects of counseling problem and counselor sex role orientation. *J Counsel Psychol* 1982;**29**:418–20.
27. Fletcher CN, Nystul MS. Attitudes towards a blind counselor and perceived effectiveness. *Psychol Rep* 1993;**73**:1091–95.
28. Garrigan JJ, Bambrick AF. Short term family therapy with emotionally disturbed children. *J Marital Fam Ther* 1975;**1**:379–85.
29. Goldfarb N. Effects of supervisory style on counselor effectiveness and facilitative responding. *J Counsel Psychol* 1978;**25**:454–60.
30. Gross WF, DeRidder LM. Significant movement in comparatively short-term counseling. *J Counsel Psychol* 1966;**13**:98–9.
31. Handley P. Relationship between supervisors' and trainees' cognitive styles and the supervision process. *J Counsel Psychol* 1982;**29**:508–15.
32. Helweg GC, Gaines LS. Subject characteristics and preferences for different approaches to psychotherapy: a multivariate study. *J Consult Clin Psychol* 1977;**45**:963–4.
33. Hill CE, Snyder JF, Schill TR. An analogue study of standard client perceptions of A and B therapists. *J Clin Psychol* 1974;**30**:94–6.
34. Hills HI, Gruszkos JR, Strong SR. Attribution and the double bind in paradoxical interventions. *Psychotherapy* 1985;**22**:779–85.
35. Hollinger-Samson N, Pearson JL. The relationship between staff empathy and depressive symptoms in nursing home residents. *Aging Ment Health* 2000;**4**:369–79.
36. Jarski RW, Gjerde CL, Bratton BD, Brown DD, Mathes SS. A comparison of four empathy instruments in simulated patient-medical student interactions. *J Med Educ* 1985;**60**:545–51.
37. Karr JT, Geist GO. Facilitation in supervision as related to facilitation in therapy. *Counsel Educ Superv* 1977;**16**:263–8.
38. Khan JA, Cross DG. Stereotypic conceptions vs perceptions of expert counsellor behaviours. *Can Counsell* 1982;**16**:201–5.
39. Kolb DL, Beutler LE, Davis CS, Crago M, Shanfield SB. Patient and therapy process variables relating to dropout and change in psychotherapy. *Psychotherapy* 1985;**22**:702–10.
40. Kulberg GE, Franco EA. Effects of A-B similarity and dissimilarity in a dyadic interaction. *Psychol Rep* 1975;**37**:1307–11.
41. LaCrosse MB. Comparative perceptions of counselor behavior: a replication and extension. *J Counsel Psychol* 1977;**24**:464–71.
42. Lanning WL. A study of the relation between group and individual counseling supervision and three relationship measures. *J Counsel Psychol* 1971;**18**:401–6.
43. Lawe CF, Horne AM, Taylor SV. Effects of pretraining procedures for clients in counseling. *Psychol Rep* 1983;**53**:327–34.
44. Lin TT. Revision and validation of the Truax–Carkhuff Relationship Questionnaire. *Measure Eval Guid* 1973;**6**:82–6.
45. Lockhart WH. Rogers' "necessary and sufficient conditions" revisited. *Br J Guid Counsell* 1984;**12**:113–23.
46. Maurer RE, Tindall JH. Effect of postural congruence on client's perception of counselor empathy. *J Counsel Psychol* 1983;**30**:158–63.
47. McNally H, Drummond R. Clients' need for social approval and perceptions of counseling relationship and outcomes. *Psychol Rep* 1973;**32**:363–6.
48. Neidigh LW. An experimental analogue examining effects of facilitative behaviours and subjects' warmth on students' perceptions of a counseling relationship. *Psychol Rep* 1991;**68**:1099–106.
49. Olson JH. Nurse-expressed empathy, patient outcomes, and development of a middle-range theory. *Image J Nurs Scholarship* 1997;**29**:71–6.
50. Olson JK. Relationships between nurse-expressed empathy, patient-perceived empathy and patient distress. *Image J Nurs Scholarship* 1995;**27**:317–22.
51. Petry RA, Thomas JR. The effect of androgyny on the quality of psychotherapeutic relationships. *Psychotherapy* 1986;**23**:249–51.
52. Pierce WD. Anxiety about the act of communicating and perceived empathy. *Psychother Theory Res Pract* 1971;**8**:120–3.
53. Reddy WB. Sensitivity training as an integral phase of counselor education. *Counsel Educ Supervis* 1970;**9**:110–15.
54. Ridge S, Campbell W, Martin D. Striving towards an understanding of conscious identification: its definition and its effects. *Counsel Psychol Q* 2002;**15**:91–105.
55. Rinaldi RC. Patient–therapist personality similarity and the therapeutic relationship. *Psychother Private Pract* 1987;**5**:11–29.
56. Roback HB, Strassberg DS. Relationship between perceived therapist-offered conditions and therapeutic movement in group psychotherapy. *Small Group Behav* 1975;**6**:345–52.
57. Salvio MA, Beutler LE, Wood JM, Engle D. The strength of the therapeutic alliance in three treatments for depression. *Psychother Res* 1992;**2**:31–6.

continued

58. Sandhu DS, Reeves TG, Portes PR. Cross-cultural counseling and neurolinguistic mirroring with Native American adolescents. *J Multicult Counsel Dev* 1993;**21**:106–18.
59. Stern MI, Bierman R. Facilitative functioning of A–B therapist types. *Psychother Theory Res Pract* 1973;**10**:44–7.
60. Vansteenwegen A. Individual and relational changes seven years after couples therapy, *Brothers, Barbara Jo (Ed)*. *Couples and change*. 1996. pp. 95–115.
61. Vansteenwegen A. Who benefits from couple therapy? A comparison of successful and failed couples. *J Sex Marital Ther* 1996;**22**:63–7.
62. Winter DA, Watson S. Personal construct psychotherapy and the cognitive therapies: different in theory but can they be differentiated in practice? *J Construct Psychol* 1999;**12**:1–22.
63. Withers LE, Wantz RA. The influence of belief systems on subjects' perceptions of empathy, warmth and genuineness. *Psychotherapy* 1993;**30**:608–15.
64. Wright W. Counselor dogmatism, willingness to disclose, and clients' empathy ratings. *J Counsel Psychol* 1975;**22**:390–4.

Unpublished references

65. Barrett–Lennard GT. Dimensions of perceived therapist response related to therapeutic change. Unpublished doctoral dissertation, University of Chicago; 1959.
66. Barrett-Lennard GT. The Relationship Inventory: Forms OS-M-64, OS-f-64 and MO-M-64 plus MO-F-64. Unpublished manuscript, University of New England; 1964.
67. Barrett-Lennard GT. Technical note on the 64-item revision of the Relationship Inventory: Forms OS-M-64, OS-f-64 and MO-M-64 plus MO-F-64. Unpublished manuscript, University of Waterloo; 1969.
68. Barrett-Lennard GT. Revision, applications and further adaptations of the Relationship Inventory. Unpublished manuscript, University of Waterloo; 1972.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C1
California Psychotherapy Alliance Scale (CALPAS) – Original

General details

Author	Marmar CR
Language	English
Country of publication/development	Canada
Publication date	NA
Publisher	NA

Purpose and overview

The CALPAS patient version (CALPAS-P) comprises five subscales: patient working capacity, patient commitment, goal disagreement, therapist negative contribution and therapist understanding and involvement. Subscales emphasise patient positive or negative contribution or therapist positive/negative contribution. All items reflect the interaction of therapist and patient. For the therapist's version (CALPAS-T), the five dimensions are patient working capacity, patient commitment, goal consensus, working strategy consensus, therapist understanding and involvement. Each of these dimensions was assessed by a single integrative judgement

Theoretical orientation	Wide range of theoretical orientations, including cognitive, behavioural, psychoanalytic, brief dynamic and person-centred therapies
Population details	Depressed older adults;[3] axis I diagnosis of depressive disorder, anxiety disorder, or combination of depressive/anxiety disorders;[2] brief psychotherapy clients[1]
Perspective	Patient, therapist and rater versions
Measure used by	Clinicians, psychotherapists, psychologists, doctoral psychologists
Other versions	24 item and 12 versions
Notes	

Areas of therapist–patient interaction addressed: Map

Therapy context: type of therapy; influence; responsibilities

Roles: expert/authority/leader

Therapist engagement: empathy; warmth; respect; sensitivity; support

Patient engagement: motivation; commitment

Framework: convergent; complementary; reciprocal; collaborative; congruent

Therapeutic techniques: responsiveness/receptiveness/attunement; exploration

Threats to the relationship: critical; defensive; resistance; withdrawal

Outcomes: compliance; satisfaction; working alliance; cohesion; expression of feelings; narrative truths; modification of working models

Therapist–client information derived from items

Dimensions

Patient version

Patient commitment	Extent to which patient values coming to treatment; optimism; willingness to continue despite moments of doubt, confusion, mistrust
Patient working capacity	Self-disclosure, self-reflection, willingness to explore own contribution to problems
Therapist understanding and involvement	Therapist empathy; respectfulness; non-judgemental acceptance; involvement in treatment process

continued

Goal disagreement	Extent to which therapist and patient are in agreement/variance concerning goals of therapy
Therapist negative contribution	Expressions of therapist annoyance/irritation/disappointment
Therapist and rater versions	
Patient working capacity	As for patient
Patient commitment	As for patient
Goal consensus	As for goal disagreement
Working strategy consensus	Extent to which patient and therapist are in agreement/at variance regarding how to proceed in therapy in order to achieve the goals
Therapist understanding and involvement	As for patient
Reliability	
The CALPAS rater version demonstrated adequate internal consistency as measured by Cronbach's alpha and adequate inter-rater reliability	
Split-half	No details
Internal consistency	0.90[1]
Inter-rater	0.94[1]
Test–retest	No details
Validity	
There is mixed evidence for the predictive and convergent validity of CALPAS. Correlations supporting the convergent validity of CALPAS with other alliance measures range from partial to adequate A five-factor solution is supported for the patient version of the CALPAS, but the intercorrelations between the therapist scales are high, meaning that a composite score is used	
Face	Dimensions derived from a variety of theoretical writings
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	CALPAS administered in the early stages of therapy was predictive of outcome rated from both patient and therapist perspectives. Significant correlations emerged between CALPAS and patient ratings of global success ($r_{22} = 0.77$, $p < 0.001$) and therapist global ratings of success CALPAS ($r_{22} = 0.55$, $p < 0.01$) CALPAS predicted outcome as measured by both the Millon Clinical Multiaxial Inventory Major Depression Scale ($r_{19} = 0.45$, $p < 0.05$) and the Beck Depression Inventory ($r_{19} = 0.45$, $p < 0.05$)[2] For the sample as a whole the therapist total alliance score was not related to outcome at the end of therapy as measured by the BDI ($r = -0.21$, ns). The relationship of therapist score and outcome was strongest in the brief dynamic therapy group ($r = -0.38$) Patient commitment was predictive of outcome at end of therapy for the sample as a whole ($r = -0.44$ ($p < 0.001$). Within the cognitive therapy group patient commitment was highly predictive of outcome ($r = -0.73$, $p < 0.01$)[3]
Construct	Construct validity was supported by the following: Patient alliance scores were more predictive of outcome than therapist composite score The relationship of therapist score and outcome was strongest in the brief dynamic therapy group (see Predictive Validity)[3]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Convergent	The CALPAS-31 correlated with Penn (0.34, ns), VTAS (0.80, $p < 0.05$), WAI-O (0.82, $p < 0.05$), WAI-C (–0.33, ns) and WAI-T (–0.22 ns)[1] The following evidence relates to the client versions of the CALPAS: The correlation between total WAI and CALPAS scores was high ($r_{22} = 0.87$, $p < 0.001$) The total CALPAS was significantly correlated with all three dimensions of the WAI: bond ($r_{22} = 0.72$, $p < 0.001$), goal ($r_{22} = 0.84$, $p < 0.001$) and task ($r_{22} = 0.79$, $p < 0.001$) The correlations between the WAI total score and CALPAS dimensions showed that all five CALPAS dimensions were significantly correlated with the WAI: patient commitment ($r_{22} = 0.85$, $p < 0.001$), patient working capacity ($r_{22} = 0.38$, $p < 0.05$), goal disagreement ($r_{22} = 0.67$, $p < 0.001$), therapist positive contribution ($r_{22} = 0.79$, $p < 0.001$) and therapist negative contribution ($r_{22} = 0.41$, $p < 0.05$)[2]
Factor structure	Intercorrelations of the five therapist scales were high (Pearson *r* values 0.73–0.87). Therefore, a composite score of CALPAS has been used to represent therapist alliance.[3] A principal components factor analysis performed on the 31-item Patient CALPAS confirmed a five-factor solution[2,3]
Responsiveness	
Discriminative (between individuals)	Global CALPAS patient scores differentiated between clients having good and poor outcomes in therapy[2] The Patient Commitment Scale on the CALPAS-P was found to be the only scale predictive of outcome, most strongly in the cognitive therapy treatment condition.[3] Therapist judgements of the alliance were not associated with differences in outcome[3]
Evaluative (within individuals across time)	No details
Number of items	31 (patient version); 5 (rater and therapist version)
Administration method	Self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1989, Williams and Wilkins
Web or scanning options	No details
Training details	Manual available: Gaston L, Marmar CR. Manual for the California Psychotherapy Alliance Scales. Unpublished manuscript. Department of Psychiatry, McGill University, Montreal; 1991
Administration/process details	Completed by therapists and clients after each therapy session
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Likert scale: five points patient's version and seven points therapist's version. Higher scores indicate better alliance
Normative data	No details

continued

Notes	
Limited therapist–patient agreement was found on this version of the CALPAS. This is consistent with earlier reports of limited patient–therapist agreement on the quality of the therapeutic relationship as assessed by the Barrett-Lennard Inventory, which suggests that alliance ratings from the two perspectives convey unique information about the treatment process The development of this original 31-item version of the CALPAS was based on an earlier version of the scales, the 41-item California Therapeutic Alliance Ratings Scales (CALTARS) intended for observer ratings. This version of the instrument is also the basis of the 24- and 12-item CALPAS instruments	
Résumé	
Strengths	Adequate internal consistency and inter-rater reliability
Weaknesses	Mixed evidence for predictive validity
	Lack of independence between the therapist scales
	At 31 items the patient version of CALPAS is quite lengthy
Areas for further research	The original CALPAS has since been developed into the 24-item version with parallel scales for patient, therapist and observer
Primary references	
1. Tichenor V, Hill CE. A comparison of six measures of working alliance. *Psychotherapy* 1989;**26**:195–9. 2. Safran JD, Wallner LK. The relative predictive validity of two therapeutic alliance measures in cognitive therapy. *Psychol Assess* 1991;**3**:188–95. 3. Marmar CR, Gaston L, Gallagher D, Thompson LW. Alliance and outcome in late-life depression. *J Nerv Ment Dis* 1989;**177**:464–72. 4. Gaston L, Marmar CR. The California Psychotherapy Alliance Scales. In Horvath AO, Greenberg LS, editors. *The working alliance: theory, research, and practice. Wiley series on personality processes*. New York: Wiley; 1994. pp. 85–108.	
Secondary reference	
5. Hatcher RL, Barends A, Hansell J, Gutfreund MJ. Patients' and therapists' shared and unique views of the therapeutic alliance: an investigation using confirmatory factor analysis in a nested design. *J Consult Clin Psychol* 1995;**63**:636–43.	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C2
California Psychotherapy Alliance Scales – Patient Version (CALPAS-P)

General details	
Authors	Marmar C, Gaston L
Language	English
Country of publication/development	Canada
Publisher	NA
Publication date	1991
Purpose and overview	
This measure assesses four theoretically derived alliance dimensions: patient commitment (PC), patient working capacity (PWC), therapist understanding and involvement (TUI) and working strategy consensus (WSC)	
Theoretical orientation	Pan-theoretical. The orientation of the practitioners in the validation study included psychodynamic, systemic, cognitive behavioural and humanistic[1]
Population details	See below
Perspective	Patient
Measure used by	Psychotherapists
Other versions	CALTARS (1984) 41 items, rater only. CALPAS (1989) patient version (31 items), rater version (five items), therapist version (five items). CALPAS-T and CALPAS-R (1988) (same family as CALPAS-P) 24 items each. CALPAS-P 12-item version.[5] The CALPAS-French translation has been used in a study of patients' views of the therapeutic alliance[8]
Notes	*Practitioners*: The practitioners participating in the validation study were psychologists (*n* = 19) and one psychiatrist, working in private practice, with an average of 9.6 years' experience in individual psychotherapy[1]
	Clients: The patient sample (*n* = 147) used to validate the measure had the following characteristics: ethnicity: all white; gender: 69% females; age: mean = 35.3 years; education: mean = 15.88 years; occupation: professionals (47%), specialised workers (25%); marital status: single (54%), married/cohabiting (29%), divorced/separated (20%), at least one child (38%); treatment history: mean = 87.3 weeks in therapy[1]
Areas of therapist–patient interaction addressed: Map	
Therapy context: type of therapy; influence; responsibilities	
Roles: expert/authority/leader	
Therapist engagement: empathy; warmth; respect; sensitivity; support	
Patient engagement: motivation; commitment	
Framework: convergent; complementary; reciprocal; collaborative; congruent	
Therapeutic techniques: responsiveness/receptiveness/attunement; exploration	
Threats to the relationship: critical; defensive; resistance; withdrawal	
Outcomes: compliance; satisfaction; working alliance; cohesion; expression of feelings; narrative truths; modification of working models	
Information derived from items	

continued

Dimensions	
Patient commitment scale (PC)	Six items. Ability to work actively and purposefully in therapy, e.g. patient's confidence that efforts will lead to change
Patient working capacity scale (PWC)	Six items. Value and investment patient places in therapy, e.g. to explore one's contribution to problems
Therapist understanding and involvement scale (TUI)	Six items. Therapist's responsiveness to client's needs, and the therapist's participation in therapy, e.g. capacity to understand the patient's point of view and suffering
Working strategy consensus (WSC)	Six items. Congruence between client and therapist on goals and strategies, e.g. how therapy should proceed
Reliability	
CALPAS Total scale demonstrated adequate reliability as estimated by Cronbach's alpha. The WSC scale demonstrated adequate reliability, the PC and TUI scales partial reliability and the PWC inadequate reliability	
Split-half	No details
Internal consistency	CALPAS-P Total scale: (0.83)
	Dimensions: PC 0.64, PWC 0.43, TUI 0.51, WSC 0.73. For PWC and TUI, one item in each was responsible for reducing the internal consistency; without this item, the Cronbach's alpha would have been 0.58; these results could not be explained by restricted variance associated with the two items[1]
	Higher internal consistency reliabilities were reported in a study using the French translation[8]
Test-retest	No details
Inter-rater	NA
Validity	
Concurrent validity with the Penn Helping Alliance Scale is generally supported: correlations of the individual scales with the Penn range from partial to adequate There is promising but mixed evidence for predictive validity CALPAS-P demonstrates discriminant validity with the Counselor Rating Form	
Face	Dimensions were derived from a variety of theoretical writings
Content	No details
Criterion (a) concurrent	Moderate to high correlations (0.37–0.60) were found between CALPAS-P scales and the Penn Helping Alliance Rating Scale (Penn HA-P).[3] High correlations were found between the total CALPAS-P score and both the Working Alliance Inventory (WAI-P) (0.83) and the Penn HA-P (0.79)[4]
Criterion (b) predictive	All CALPAS-P scales correlated positively with patient satisfaction as measured by CSQ-8; PC 0.43, PWC 0.39, TUI 0.65, WSC 0.65 and Total score 0.66, all at $p < 0.0035$. Findings suggested a potentially greater role of the working alliance (PWC) in dynamic psychotherapy (as compared with cognitive-behavioural) in correlating with treatment satisfaction[1]
	In a sample of elderly depressed patients CALPAS-P scores contributed to large amounts of outcome variance (as measured by the BDI), over and above initial symptomatology and in-treatment symptom change, at the fifth, tenth and 15th sessions, but the results were not statistically significant[7]
Construct	See Convergent and Discriminant validity
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Convergent	Negative correlations were found between three dimensions of CALPAS-P and patients' symptomatology (as measured by SCL-10); PC –0.43; PWC –0.26; WSC –0.28 ($p < 0.0035$), indicating that greater symptomatology diminished the patients' capacity to become engaged in therapy, to work in therapy, and to have a sense of working with the therapist towards agreed-on goals. These same dimensions of CALPAS-P also exhibited negative correlations with patients' intimacy as measured by the IIP intimacy subscale (PC –0.26; PWC –0.30; WSC –0.25, $p < 0.0035$) suggesting that patients with greater intimacy difficulties reported poorer therapeutic and working alliance[1]
Discriminant	To show that the alliance as measured by the CALPAS differs from other related constructs, an exploratory factor analysis with oblique rotation was conducted. The included variables were the four CALPAS-P scales; and three subscales of expertness, attractiveness and trustworthiness from the Counselor Rating Form. Two factors emerged: an alliance factor composed of the four CALPAS-P scales, and a perceived therapist influence factor composed of the other three scales, thus discriminating the CALPAS-P from the other related constructs[3]
Factor structure	Correlations between the four CALPAS-P scales ranged from 0.37 to 0.62, indicating 14–38% of shared variance. The highest common variance was shared by the WSC and TUI scales, which is consistent with previous research indicating that scales reflecting patient–therapist agreement on goals and tasks correlate highly with other alliance dimensions. Correlations between each scale and the total score ranged from 0.73 to 0.82.[1] Similar correlations between CALPAS-P scales were found as in refs 1 and 3. Large correlations between CALPAS-P scales were found for the alliance within the most recent session (0.59–0.81) and for the psychotherapy received so far (0.22–0.81).[4] Confirmatory factor analysis was used to examine the factor structure of the CALPAS-P. A single-factor model was compared with a bilevel model, where four alliance factors were embedded within a general alliance factor. The bilevel model was the best fit to the data, although it was not a good fit. When the CALPAS-P was reduced to the 12 most discriminative items of the four alliance dimensions, the bilevel model did appear to be a good fit, supporting the theoretical model on which the CALPAS was elaborated[5]
Responsiveness	
Discriminative (between individuals)	In two studies, CALPAS-P levels were not found to be different in behavioural, cognitive, humanistic and/or dynamic psychotherapy.[1,7] No association was found between CALPAS-P scores and patients' age, level of education, yearly income, gender, marital status, number of sessions in therapy or social desirability (as measured by the Marlowe–Crowne Social Desirability Scale). No association was found between CALPAS-P scores and therapists' years of practice, gender or theoretical orientation[3]
Evaluative (within individuals across time)	No changes in CALPAS-P level over time were observed in a study of depressed elderly patients. Substantial increases in outcome variance were found to be accounted for by CALPAS-P scores from early and late sessions, but the results were not significant owing to the limited power of the study. Ratings of positive transference were associated with outcome, only at the end of therapy.[7] In another study of the alliance over time, no significant change in CALPAS-P scores was found from the fifth to the tenth session[8]
Acceptability	
Number of items	24
Administration method	Questionnaire
Time taken to complete	10–15 minutes
Flesch reading age	No details
Translations	French
Access by ethnic minorities	No details

continued

Feasibility	
Copyright	1991, American Psychological Association
Web or scanning options	No details
Training details	Manual available from Dr Louise Gaston[6]
Administration process details	The patient answers the CALPAS-P right after the completion of a therapy session. The patient indicates the degree to which each statement describes their experience during the session. Information provided by patients is confidential, but patients are welcome to discuss any item or reaction to an item with their therapist
Support from measure developers	Information on CALPAS available from Dr Louise Gaston
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert. There are six items for each dimension, making a total of 24 items. For each item the patient indicates the degree to which it described his/her experience in the therapy session just completed. Each item is scored on a seven-point Likert scale rated from 1 (not at all) to 7 (very much so). Half the items are positively phrased and half negatively phrased to counter the tendency of providing positive responses
Normative data	No details
Notes	
CALPAS-P was developed from the CALPAS 31-item patient version (1989), which in turn was developed from CALTARS (1984). Some items in CALPAS-P were derived from the original scales (Marmar, Gaston *et al*., 1989; Marmar, Weiss *et al*., 1989); others were developed to take into account theoretical issues omitted in previous alliance measures. CALPAS-P has been adapted for use in group psychotherapy, but little information is available[2]	
Résumé	
Strengths	Findings indicate that the CALPAS-P is a promising measure for assessing dimensions of the alliance. Internal consistency for the whole CALPAS-P was high. The four CALPAS-P scales were moderately intercorrelated, suggesting that they reflect independent dimensions of the alliance. CALPAS-P scales yielded large correlations to criterion variables and were not related to social desirability. CALPAS-P scales did not vary across therapy modalities or number of sessions, suggesting that the measure may be amenable to assess the alliance across therapy modalities and at different points in therapy[1]
Weaknesses	Low alpha coefficients were found with two CALPAS-P scales (PWC and TUI). Whether these estimates are adequate is debatable[1]
Areas for further research	Further confirmatory factor analysis is needed to test the theoretical assumptions underlying the structure of the CALPAS-P (24 item). Future work on the criterion-related validity of the CALPAS-P is required, e.g. to relate CALPAS-P scales to therapist-completed outcome measures or observers, to substantiate its predictive validity[1,2]
Primary references	
1. Gaston L. Reliability and criterion-related validity of the California Psychotherapy Alliance Scales – Patient version. *Psychol Assess* 1991;**3**:68–74. 2. Gaston L, Marmar CR. The California Psychotherapy Alliance Scales. In Horvath AO, Greenberg LS, editors. *The working alliance: theory, research, and practice. Wiley series on personality processes*. New York: Wiley; 1994. pp. 85–108. 3. Sabourin S, Coallier JC, Cournoyer LG, Gaston L. Further aspects of the validity of the California Psychotherapy Alliance Scales. Paper presented at the Meeting of the Society for Psychotherapy Research, Wintergreen, VA, June 1990. 4. Hatcher R, Hansell J, Barends A, Leary K, Stuart J, White K. Comparison of several psychotherapy alliance measures. Paper presented at the Meeting of the Society for Psychotherapy Research, Wintergreen, VA, June 1990.	
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

5. Gaston L, Sabourin S, Hatcher R, Hansell J. Confirmatory factor analysis of the CALPAS-P and its short version. Paper presented at the meeting of the International Society for Psychotherapy Research, Berkeley, CA, June 1992.
6. Gaston L, Marmar CR. Manual for the California Psychotherapy Alliance Scales – CALPAS. Unpublished manuscript, Department of Psychiatry, McGill University, Montreal, 1991.

Secondary references

7. Gaston L, Marmar CR, Thompson LW, Gallagher D. Alliance prediction of outcome beyond in-treatment symptomatic change as psychotherapy progresses. *Psychother Res* 1991;**1**:104–13.
8. Bachelor A, Salame R. Participants' perceptions of dimensions of the therapeutic alliance over the course of therapy. *J Psychother Pract Res* 2000;**9**:39–53.

C3
California Psychotherapy Alliance Scales – Rater Version (CALPAS-R)

General details	
Authors	Marmar C, Gaston L
Language	English
Country of publication/development	Canada
Publication date	1991
Publisher	NA
Purpose and overview	
This measure assesses four theoretically derived alliance dimensions: patient commitment (PC), patient working capacity (PWC), therapist understanding and involvement (TUI) and working strategy consensus (WSC)	
Theoretical orientation	Various/range
Population details	Therapists
Perspective	Rater
Measure used by	Psychotherapists
Other versions	CALPAS (1989) patient version (31 items), rater version (five items), therapist version (five items). CALPAS-T and CALPAS-R (1988) (same family as CALPAS-P) 24 items each. CALPAS-P 12-item version CALTARS (1984) 41 items, rater only
Notes	
Areas of therapist–patient interaction addressed: Map	
Therapy context: type of therapy; influence; responsibilities	
Roles: expert/authority/leader	
Therapist engagement: empathy; warmth; respect; sensitivity; support	
Patient engagement: motivation; commitment	
Framework: convergent; complementary; reciprocal; collaborative; congruent	
Therapeutic techniques: responsiveness/receptiveness/attunement; exploration	
Threats to the relationship: critical; defensive; resistance; withdrawal	
Outcomes: compliance; satisfaction; working alliance; cohesion; expression of feelings; narrative truths; modification of working models	
Information derived from items	
Dimensions	
Patient commitment scale (PC)	Six items. Ability to work actively and purposefully in therapy, e.g. patient's confidence that efforts will lead to change
Patient working capacity scale (PWC)	Six items. Value and investment patient places in therapy, e.g. to explore one's contribution to problems
Therapist understanding and involvement scale (TUI)	Six items. Therapist's responsiveness to client's needs, and the therapist's participation in therapy, e.g. capacity to understand the patient's point of view and suffering
Working strategy consensus (WSC)	Six items. Congruence between client and therapist on goals and strategies, e.g. how therapy should proceed

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
The CALPAS-R has adequate internal consistency as estimated by Cronbach's alpha and generally adequate inter-rater reliability when more than two raters are employed	
Split-half	No details
Internal consistency	PWC-positive 0.95, PWC-negative 0.87, PC 0.94, WSC 0.95, TUI 0.93[1] Total scale: 0.92, PWC 0.92, PC 0.88, WSC 0.83, TUI 0.85[2] PWC: 0.95, PC 0.96, WSC 0.95, TUI 0.97[4,6]
Test–retest	No details
Inter-rater	ICCs for three raters: PWC-positive 0.90, PWC-negative 0.82, PC 0.91, WSC 0.86, TUI 0.89[1] PWC 0.76, PC 0.82, WSC 0.77, 0.63[7] ICCs for two raters: Total scale 0.75[2] Total scale 0.76[3] PWC 0.64, PC 0.68, WSC 0.69, TUI 0.52[7] Total scale 0.90 Four scales range 0.83 to 0.92[8] ICCs for two teams of raters: PWC 0.94, PC 0.94, WSC 0.89, TUI 0.97[4,6] ICCs for 1 rating: PWC 0.52, PC 0.60, WSC 0.53, TUI 0.36

Validity	
CALPAS-R generally demonstrates adequate convergent, construct and predictive validity across studies	
Face	Dimensions were derived from a variety of theoretical writings on the working alliance[4,5]
Content	No details
Criterion (a) concurrent	See Convergent
Criterion (b) predictive	Correlations of CALPAS with outcome as measured by days abstinent from cocaine:[3] All treatments 0.37**, cognitive behavioural therapy 0.56**, twelve-step facilitation therapy 0.28[3] ** $p < 0.001$ CALPAS significantly correlated with target objectives as assessed by patients at the end of short-term therapy (–0.44, $p = 0.09$), but not significantly with objectives as assessed by an independent evaluation (–0.10, $p > 0.05$)[4,6] In short-term therapy PWC and PC ratings accounted for 13%* and 28%* respectively of symptomatology variance, but did not predict interpersonal problems[6] In long-term therapy only PWC ratings predicted variance in outcome: 25%* and 30%* for symptomatology and interpersonal problems, respectively[6] * $p < 0.05$[6] Across therapies PWC ratings predicted variance in outcome on the BDI and Hamilton Rating Scale for Depression (HRSD) at 8%* and 7%*, respectively[7] Across therapies PC ratings predicted variance in outcome on the BDI at 15%*[7] When analyses were performed separately for each treatment, PC ratings only predicted variance in the cognitive therapy condition at 15%*[7] * $p < 0.05$[7] No significant associations between CALPAS at session 5 and outcome[8]

continued

Construct	Hypothesis that ratings of alliance would predict outcome variance was partially supported. See predictive validity section under 6 and 7
Convergent	Correlations of mean ratings of CALPAS scales with Session Evaluation Questionnaire (SEQ): PWC-positive and SEQ Depth 0.70* PWC-negative and SEQ Depth: –0.49 WSC and SEQ Depth: 0.68* PC and SEQ Depth: 0.63* TUI and SEQ Depth: 0.49 PWC-positive and SEQ helpfulness: 0.46 PC and SEQ helpfulness: 0.57 WSC and SEQ helpfulness: 0.47 TUI and SEQ helpfulness: 0.58*[1] * $p < 0.05$ Correlation of CALPAS Total scale with Penn Helping Alliance Total scale (Penn): (0.54, $p < 0.001$) and Vanderbilt Therapeutic Alliance Scale (VTAS): (0.60, $p < 0.001$) (see ref. 2 for further intercorrelations between dimensions of measures) CALPAS and Penn: 0.62** CALPAS and VTAS: 0.38* CALPAS and Working Alliance Inventory – Observer (WAI-O): 0.37 CALPAS and Working Alliance Inventory – Client (WAI-C): 0.31 CALPAS and Working Alliance Inventory – Therapist (WAI-T): 0.51*[3] * $p < 0.01$; ** $p < 0.001$[3]
Discriminant	TUI and SEQ smoothness –0.39[1]
Factor structure	The CALPAS-R comprises the four dimensions described above. A factorial analysis of the patient version of the CALPAS provides support for this factor structure[7] Significant intercorrelations between dimensions suggest they are empirically non-independent of each other[1] Correlations among CALPAS scales ranged from 0.33 to 0.83. The lowest correlation was observed between the TUI scale and the other scales[4,6] The four scales correlated substantially with each other and with the total alliance score. Relatively small intercorrelations between the TUI and the PC and PWC scales
Responsiveness	
Discriminative (between individuals)	No significant differences between CALPAS ratings for 8- and 16-session treatments[1] Better working alliances were observed in more introspective types of treatment such as cognitive and brief dynamic therapy[7]
Evaluative (within individuals across time)	An increase in therapeutic alliance was observed over time[7]
Acceptability	
Number of Items	24
Administration method	Rater-completed questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	French
Access by ethnic minorities	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Feasibility	
Copyright	1991, American Psychological Association
Web or scanning options	No details
Training details	Full training details and rating protocol for judges are contained in manual available from Dr Louise Gaston. It is recommended that clinical judges have several years of experience and training in psychotherapy[5]
Administration process details	For each therapy session the rating proceeds in two steps. First, while reviewing a therapy session, the clinical judges take note of their observations related to each element of the alliance. After reviewing the session, the clinical judges assess the degree to which each subcomponent has occurred during the therapy session and indicate their judgements on the seven-point scale
Support from measure developers	Information on CALPAS-R available from Dr Louise Gaston
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert. There are six items for each dimension, making a total of 24 items. For each item the patient indicates the degree to which it described his/her experience in the therapy session just completed. Each item is scored on a seven-point Likert scale rated from 1 (not at all) to 7 (very much so). Half the items are positively phrased and half negatively phrased to counter the tendency of providing positive responses
Normative data	No details
Notes	
CALPAS-R 24-item version was developed in parallel with the CALPAS-P 24-item version. CALPAS-R differs from the original version of the rater instrument which comprises five items	
Résumé	
Strengths	CALPAS-R 24-item version is a development of the original five-item version of the instrument and allows therapists greater discrimination in their judgement of the alliance. Also, the intercorrelations between the five items in the original version of the CALPAS were found to be too high Extensive evidence and support for reliability and validity
Weaknesses	Mixed evidence for predictive validity
Areas for further research	Further testing of psychometric properties in more culturally/ethnically diverse samples

Primary references

1. Barkham M, Agnew RM, Culverwell A. The California Psychotherapy Alliance Scales: a pilot study of dimensions and elements. *Br J Med Psychol* 1993;**66**:157–65.
2. Cecero JJ, Fenton LR, Frankforter TL, Nich C, Carroll KM. Focus on therapeutic alliance: the psychometric properties of six measures across three treatments. *Psychotherapy* 2001;**38**:1–11.
3. Fenton LR, Cecero JJ, Nich C, Frankforter TL, Carroll KM. Perspective is everything: the predictive validity working alliance instruments. *J Psychother Pract Res* 2001;**10**:262–8.
4. Gaston L, Marmar CR. The California Psychotherapy Alliance Scales. In Horvath AO, Greenberg, LS, editors. *The working alliance: theory, research, and practice. Wiley series on personality processes*. New York: Wiley; 1994. pp. 85–108.
5. Gaston L, Marmar CR. Manual for the California Psychotherapy Alliance Scales. Unpublished manuscript. Department of Psychiatry, McGill University, Montreal; 1991.

Secondary references

6. Gaston L, Piper WE, Debbane EG, Bienvenu J.-P, Garant J. Alliance and technique for predicting outcome in short- and long-term analytic psychotherapy. *Psychother Res* 1994;**4**:121–35.
7. Gaston L, Thompson L, Gallagher D, Cournoyer L.-G, Gagnon R. Alliance, technique, and their interactions in predicting outcome of behavioral, cognitive, and brief dynamic therapy. *Psychother Res* 1998;**8**:190–209.
8. Price PB, Jones EE. Examining the alliance using the Psychotherapy Process Q-Set. *Psychotherapy* 1998;**35**:392–404.

C4
California Psychotherapy Alliance Scales – Therapist Version (CALPAS-T)

General details	
Authors	Marmar C, Gaston L
Language	English
Country of publication/development	Canada
Publication date	1991
Publisher	NA
Purpose and overview	
This measure assesses four theoretically derived alliance dimensions: patient commitment (PC), patient working capacity (PWC), therapist understanding and involvement (TUI) and working strategy consensus (WSC)	
Theoretical orientation	Various/range
Population details	Therapists
Perspective	Therapist
Measure used by	Psychotherapists
Other versions	CALTARS (1984) 41 items, rater only, CALPAS (1989) patient version (31 items), rater version (five items), therapist version (five items), CALPAS-T and CALPAS-R (1988) (same family as CALPAS-P) 24 items each. CALPAS-P 12-item version
Notes	
Areas of therapist–patient interaction addressed: Map	
Therapy context: type of therapy; influence; responsibilities	
Roles: expert/authority/leader	
Therapist engagement: empathy; warmth; respect; sensitivity; support	
Patient engagement: motivation; commitment	
Framework: convergent; complementary; reciprocal; collaborative; congruent	
Therapeutic techniques: responsiveness/receptiveness/attunement; exploration	
Threats to the relationship: critical; defensive; resistance; withdrawal	
Outcomes: compliance; satisfaction; working alliance; cohesion; expression of feelings; narrative truths; modification of working models	
Information derived from items	
Dimensions	
Patient commitment scale (PC)	Six items. Ability to work actively and purposefully in therapy, e.g. patient's confidence that efforts will lead to change
Patient working capacity scale (PWC)	Six items. Value and investment patient places in therapy, e.g. to explore one's contribution to problems
Therapist understanding and involvement scale (TUI)	Six items. Therapist's responsiveness to client's needs, and the therapist's participation in therapy, e.g. capacity to understand the patient's point of view and sufferings
Working strategy consensus (WSC)	Six items. Congruence between client and therapist on goals and strategies, e.g. how therapy should proceed
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
Examination of the internal structure of CALPAS-T has not yet been conducted[2]	
Split-half	No details
Internal consistency	No details
Test–retest	No details
Inter-rater	NA
Validity	
The CALPAS-T demonstrates adequate convergent validity with the therapist versions of the WAI and the Penn Helping Alliance Rating Scales	
The CALPAS-T does not have evidence regarding its factor structure, but consists of the same four scales as the CALPAS-P	
Face	Dimensions were derived from a variety of theoretical writings
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	CALPAS-T Total score correlated highly with the Working Alliance Inventory – Therapist (WAI-T) (0.79) and the Penn Helping Alliance Rating Scale – Therapist (Penn-T) – (0.71)[3]
Discriminant	No details
Factor structure	The CALPAS-T consists of the same four scales as the CALPAS-P, but there are currently no details on the internal structure of the therapist version of the instrument
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individuals across time)	No details
Acceptability	
Number of items	24
Administration method	Self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	French
Access by ethnic minorities	No details
Feasibility	
Copyright	1991, American Psychological Association
Web or scanning options	No details
Training details	Full training details and rating protocol for therapists is contained in a manual available from Dr Louise Gaston[1]
Administration process details	After the completion of a therapy session, a therapist proceeds to rate the CALPAS-T by indicating on a seven-point scale the degree to which the phenomena described by an item had occurred in the session
Support from measure developers	Information on CALPAS-T available from Dr Louise Gaston
FAQ facility	No details

continued

Precision	
Scale type	Ordinal, Likert. There are six items for each dimension, making a total of 24 items. For each item the patient indicates the degree to which it described his/her experience in the therapy session just completed. Each item is scored on a seven-point Likert scale rated from 1 (not at all) to 7 (very much so). Half the items are positively phrased and half negatively phrased to counter the tendency of providing positive responses
Normative data	No details
Notes	
CALPAS-T 24-item version was developed in parallel with the CALPAS-P 24-item version. CALPAS-T differs from the original version of the therapist instrument, which comprises five items	
Résumé	
Strengths	CALPAS-T 24-item version, in contrast to the original five-item version of the instrument, allows therapists greater discrimination in their judgement of the alliance
Weaknesses	Little psychometric information
Areas for further research	Further testing of psychometric properties

Primary references

1. Gaston L, Marmar CR. *Manual for the California Psychotherapy Alliance Scales*. Unpublished manuscript, Department of Psychiatry, McGill University, Montreal; 1991.
2. Gaston L, Marmar CR. The California Psychotherapy Alliance Scales. In Horvath AO, Greenberg LS, editors. *The working alliance: theory, research, and practice. Wiley series on personality processes*. New York: Wiley; 1994. pp. 85–108.
3. Hatcher R, Hansell J, Barends A, Leary K, Stuart J, White K. *Comparison of several psychotherapy alliance measures*. Paper presented at the Meeting of the Society for Psychotherapy Research, Wintergreen, VA; June 1990.

Secondary reference

4. Bachelor A, Salame R. Participants' perceptions of dimensions of the therapeutic alliance over the course of therapy. *J Psychother Pract Res* 2000;**9**:39–53.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C5
California Therapeutic Alliance Rating System (CALTARS)

General details	
Author	Marmar C
Language	English
Country of publication/development	USA
Publication date	1981
Publisher	NA
Purpose and overview	
The CALTARS measures therapist and patient positive and negative contributions to the alliance	
Theoretical orientation	Psychoanalytic, but not intended for exclusive use in psychoanalytic therapy
Population details	Adults with neurotic level reactions to traumatic life events[3] and adjustment disorders[1,2]
Perspective	Observer rated
Measure used by	Psychotherapists, researchers
Other versions	A modified final 42-item version of CALTARS rated by judges, patients and therapists, consisting of 21 therapist contribution items (11 positive, ten negative) and 21 patient contribution items (11 positive, ten negative)
Notes	
Areas of therapist–patient interaction addressed: Map	
Therapy context: power/coercion; responsibilities	
Therapist engagement: empathy/sensitivity; warmth; respect; support; listening; hope/encouragement; praise/affirmation	
Patient engagement	
Framework	
Therapeutic techniques	
Information derived from example items	
Dimensions	
	Examples
Patient positive contribution (PPC)	11 items. The patient indicates that he/she experiences the therapist as understanding and accepting
Patient negative contribution (PNC)	Ten items. The patient acts in a hostile, attacking and critical manner towards the therapist
Therapist positive contribution (TPC)	11 items. The therapist is hopeful and encouraging, conveying the belief that the patient has made, is making, or can make progress
Therapist negative contribution (TNC)	Ten items. The therapist criticises the patient and/or behaves in such a way that the patient feels put down
Reliability	
The CALTARS demonstrates adequate internal consistency and partial/adequate inter-rater reliability across studies	
Split-half	No details
Internal consistency	Alpha coefficients for each of the items ranged from 0.82 to 0.85[1]
	All alpha coefficients were between 0.76 and 0.94[2]
	Alpha coefficient: therapist total contribution 0.88, patient total contribution 0.94[3]
	continued

Inter-rater	*ICCs (two judges)* Two judges: coefficients for the scales ranged from 0.82 to 0.85 Seven judges in pairings: TPC 0.75, TNC 0.69, PPC 0.76, PNC 0.65[1] Reliabilities of the five scales were assessed at session and treatment levels. Two judges were used: Session-level reliabilities were marginally acceptable or low ($r = 0.19$ to 0.62). At treatment level the reliabilities were acceptable ($r = 0.76$ to 0.81)[2] Finn's *r* statistic used to calculate inter-rater reliability: Therapist items over 40 judged therapist hours: mean 0.82 and median 0.85 Patient items over 40 judged therapist hours: mean 0.76 and median 0.78[3]
Test–retest	No details
Validity	
Extensive validity evidence is available on the CALTARS. CALTARS demonstrates partial/adequate convergent, concurrent and predictive validity across studies and with some scales Adequate discriminant validity with the SCL-90 has been demonstrated by CALTARS, with the exception of the patient working capacity scale A four-factor structure supporting the four theoretically derived scales listed in dimensions has been supported by factor analytic studies. However, a five-factor solution of patient commitment, patient working capacity, patient hostile resistance, therapist negative contribution and therapist understanding and involvement provides a clearer conceptualisation of the theoretical constructs and forms the basis for the development of the California Psychotherapy Alliance Scales (CALPAS)	
Face	The items of CALTARS were generated from items selected from the scales of Luborsky, Hartley and Strupp, and Gomes-Schwartz, as well as new items generated from intensive case studies. Items involving action, technique or specific response were deleted
Content	See above
Criterion (a) concurrent	The results of the principal components analysis are consistent with the factor structure of the Vanderbilt Therapeutic Alliance Scale (Hartley and Strupp, 1983). They found six factors, five similar to this study's, plus an unrelated anxiety Several concurrent assessments were conducted with encouraging results. Patient hostile resistance (PHR), patient commitment (PC), patient working capacity (PWC) and therapist understanding and involvement (TUI) were all related, as expected, to the therapist's action scale (TAS) and Patient Experiencing Scale (PES) PHR and TAS (discuss avoidance): $r = 0.31$* PHR and TAS (address view of therapist): $r = 0.33$* PC and TAS (link reaction towards therapist to parents): $r = -0.50$*** TNC and TAS (discuss avoidance): $r = 0.31$* PWC and TAS (address view of therapist): $r = -0.30$* PWC and TAS (link reaction towards therapist to parents): $r = -0.50$*** TUI and PES (modal experiencing): $r = 0.42$** TUI and PES (peak experiencing): $r = 0.41$** PC and PES (modal experiencing): $r = 0.41$** PC and PES (peak experiencing) $r = 0.37$** PWC and PES (modal experiencing) $r = 0.50$*** PWC and PES (peak experiencing) $r = 0.50$*** *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$[2]
Criterion (b) predictive	Partial correlations were computed between the mean score, over four sessions, for each subscale and outcome. Only one of the eight coefficients was significant: PNC was significantly negatively related to the rate of symptom decline[1]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	Multiple regression assessed the effect on outcome of the interaction of the scales with patient motivation and developmental level of self-concept. Therapist contributions still did not predict outcome. Motivation interacted with patient contributions to predict outcome meaningfully. Motivation accounted for an additional 10% of the variance in outcome (increment in $R^2 = 0.10$, $p < 0.05$)[1]
	Partial correlations showed that the PWC scale predicted outcome. Scale scores were associated with greater symptom improvement (partial $r_{48} = -0.29$, $p < 0.05$.) and interpersonal functioning (partial $r_{45} = 0.39$, $p < 0.01$)[2]
Construct	See under Convergent, Predictive and Discriminant validity sections. Hypotheses relating to these components of validity were generated and tested[1,2]
Convergent	Based on ratings using the Patterns of Individual Change Scales and ratings on a measure of developmental self-concept, two estimates of patient pretreatment relationship and stability and social functioning were made and correlated with the scales. Of eight correlations, one was significant. Pretherapy developmental level was associated with the PPC scale ($r = 0.40$, $p < 0.01$)[1]
	The alliance scales were correlated with the following patient pretreatment characteristics: TUI and patient's educational level ($r = 0.037$, $p < 0.05$). PC and patient's educational level ($r = 0.58$, $p < 0.001$) Life events questionnaire and PC ($r = -0.42$, $p < 0.01$) Life events questionnaire and PWC ($r = -0.50$, $p < 0.001$) PWC and motivation ($r = 0.48$, $p < 0.001$) PWC and relationship composite (from Patterns of Individual Change scale): ($r = 0.33$, $p < 0.05$) PWC and symptom checklist (SCL-90) ($r = -0.29$, $p < 0.05$)[2]
Discriminant	The scale's four subscale mean ratings over four sessions were correlated with patient pretherapy score on the SCL-90 (self-report symptom checklist). There were no significant correlations ($p > 0.05$), meaning that the scales do not measure patient symptomatic distress[1]
	Four of the five alliance scale scores were uncorrelated with initial SCL-90 symptomatology scores. Patient Working Capacity showed a moderate negative association with initial SCL-90 symptomatology ($r = -0.29$, $p < 0.05$)[2]
Factor structure	The four scales listed under dimensions were supported by factor analytic findings. All coefficients of the four scales were significant and ranged from 0.46 to 0.81[1]
	Principal components analysis yielded five components accounting in total for 63% of the variance. Five component-based scales were constructed by selecting items that were conceptually relevant and loaded 0.52 or more on that component and not more than 0.35 on any other component. Nine items were excluded. To assess whether the component-based scales represented the original component solution, correlations were computed between the original components and the corresponding scales. Coefficients were 0.71, 0.94, 0.94, 0.95 and 0.97. Intercorrelations between the five component-based scales were as expected, e.g. patient hostile resistance was negatively related to patient working capacity[2]

continued

Responsiveness	
Discriminative (between individuals)	See under Convergent validity[1,2]
	Mann Whitney *U* test:
	Patient total contribution scale discriminated between good and poor outcome groups ($U = 0$, $p = 0.01$)
	Therapist total contribution did not discriminate between the two outcome groups ($U = 7$, $p = 0.31$)[3]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	41
Administration method	Judge-completed rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	French
Access by ethnic minorities	No details
Feasibility	
Copyright	1981, American Psychiatric Association
Web or scanning options	No details
Training details	12 hours' training. Recommended that experienced clinicians be used as raters as they can be more readily trained to a criterion of reliability (0.70)
	A manual providing an operational definition for each item was generated during the development of the scales
Administration/process details	In the case of a 12-session time-limited psychodynamic therapy, sessions 2, 5, 8 and 11 have been sampled. Recommended that the first, middle or last 20 minutes of each session is chosen on a random basis for rating. Segments coded and randomly presented to raters to avoid generalisation/halo effect. Raters should be blind to patient outcomes
	Recommended that videotape-recordings are used
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Each of the items is rated on a five-point Likert-type scale of intensity for presence ranging from 0 (not present) to 5 (intensely present)
Normative data	No details
Notes	
All validity and reliability assessments relating to ref. 2 were conducted with the five component-based scales produced from the principal components analysis. These five scales are: therapist understanding and involvement, patient hostile resistance, patient commitment, therapist negative contribution and patient working capacity	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Résumé	
Strengths	Extensive research evidence indicates that the CALTARS measures alliance in a clinically meaningful and psychometrically robust manner
	Proven link with outcome; useful for process outcome research
Weaknesses	Highly experienced clinicians recommended as raters. Advanced raters may be used, but require at least 12 hours of intensive training. Therefore, the CALTARS would not be recommended for routine use in service settings
Areas for further research	CALTARS has been modified and developed to produce the CALPAS scales

Primary references

1. Marmar CR, Horowitz MJ, Weiss DS, Marziali E. The development of the Therapeutic Alliance Rating System. In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford; 1986. pp. 367–90
2. Marmar CR, Weiss DS, Gaston L. Toward the validation of the California Therapeutic Alliance Rating System. *Psychol Assess* 1989;**1**:46–52.
3. Marziali E, Marmar C, Krupnick J. Therapeutic alliance scales: development and relationship to psychotherapy outcome. *Am J Psychiatry* 1981;**138**:361–4.

Secondary references

4. Eaton TT, Abeles N, Gutfreund MJ. Negative indicators, therapeutic alliance, and therapy outcome. *Psychother Res* 1993;**3**:115–23.
5. Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook*. New York: Guilford Press; 1986.
6. Hentschel U, Bijleveld CCJH. It takes two to do therapy: On differential aspects in the formation of therapeutic alliance. *Psychother Res* 1995;**5**:22–32.

C6
California Therapeutic Alliance Rating System Scales (CALTARS Scales)

General details	
Author	Marmar C
Language	English
Country of publication/development	USA
Publication date	1984
Publisher	NA
Purpose and overview	
A development of the Therapeutic Alliance Rating System: patient-rated and therapist-rated scales were developed to parallel the dimensions included in the original judge-rated scale. The original rating system was also in part reorganised. Some items were reformulated to ensure unidimensionality, uniformity of language and a balance of positive and negative items. The CALTARS scales measure therapist and patient positive and negative contributions to the alliance from each of the three perspectives of observer, therapist and patient	
Theoretical orientation	Psychoanalytic, but not intended for exclusive use in psychoanalytic therapy
Population details	Adults with neuroses from psychiatric outpatients service;[2] female undergraduates with psychoneuroses, interpersonal problems and personality disorders
Perspective	Observer, client and therapist rated
Measure used by	Psychotherapists, researchers
Other versions	The judge-, therapist- and patient-rated scales are a development of the original 41-item judge-rated system, CALTARS
Notes	
Areas of therapist–patient interaction addressed: Map	
Therapy context: power/coercion; responsibilities	
Therapist engagement: empathy/sensitivity; warmth; respect; support; listening; hope/encouragement; praise/affirmation	
Patient engagement	
Framework	
Therapeutic techniques	
Information derived from example items	
Dimensions	
	Examples
Patient positive contribution (PPC)	11 items. The patient indicates that he/she experiences the therapist as understanding and accepting
Patient negative contribution (PNC)	Ten items. The patient acts in a hostile, attacking and critical manner towards the therapist
Therapist positive contribution (TPC)	11 items. The therapist is hopeful and encouraging, conveying the belief that the patient has made is making or can make progress
Therapist negative contribution (TNC)	Ten items. The therapist criticises the patient and/or behaves in such a way that the patient feels put down

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
The CALTARS alliance scales demonstrate adequate internal consistency and partial to adequate inter-rater reliability across scales	
Split-half	No details
Internal consistency	Cronbach's alpha coefficients ranged from 0.81 to 0.93[1,2]
Inter-rater	*ICCs* Patient contribution positive and negative items: range 0.60 to 0.83 Therapist contribution positive and negative items: range 0.61 to 0.77[1,2]
Test–retest	No details
Validity	
There is mixed supportive evidence for predictive validity. The scales demonstrated partial convergent validity with patients' pretherapy measures of social adjustment and symptomatic status. A two-factor (positive and negative items) solution was supported	
Face	The items of CALTARS scales were taken from the original CALTARS rating system. Some items were reformulated to ensure unidimensionality, uniformity of language, and a balance of positive and negative items
Content	See above
Criterion (a) concurrent	No details
Criterion (b) predictive	For each of the three rating systems, the mean scores of the separate patients' and therapists' negative contributions were used in computations of partial correlations with six outcome measures. 72 partial correlations were computed. Of these, 27 coefficients were significant (see Table 10.2 in ref. 1 for details). The range of the significant correlation coefficients across outcome measures is as follows. Consistently, the patients' positive and negative contributions in each of the three rating systems were the best predictors of outcome. All correlations were in the expected direction *Patient-rated scale* PPOS: range –0.34 to 0.57 PNEG: range –0.31 to –0.27 TPOS: range –0.30 to 0.47 TNEG: No significant correlations *Therapist-rated scale* PPOS: range –0.37 to 0.52 PNEG: range –0.38 to –0.35 TPOS: range –0.30 to 0.32 TNEG: –0.29 *Judge-rated scale* PPOS: range 0.25 to 0.59 PNEG: range –0.43 to –0.27 TPOS: 0.30 TNEG: no significant correlations Above information taken from refs 1 and 2
Construct	No details
Convergent	The total mean scores of the six sessions' alliance ratings of the patients' positive and negative contributions within each of the three ratings systems were correlated with three pretherapy measures of the patients' social adjustment and symptomatic status Patients' pretherapy ratings of social adjustment correlated significantly with: patient negative self-ratings and ratings of therapist negative contributions therapist's ratings of patients positive and negative contributions external judges' ratings of the patients' negative contributions patients' judgements of the therapists' negativeness was significantly correlated with both the pretherapy symptom index and mood scores at *r* values of 0.26 and 0.25, respectively Above information taken from refs 1 and 2
	continued

Discriminant	No details
Factor structure	A principal components factor analysis was carried out on each of the alliance measurement systems: patient rated, therapist rated and external judge rated. The factor structure was consistent for all three measurement systems across all six sessions and for each of the subscales within each measurement system. Two factors emerged. The first factor consisted of all the positive items. The second factor consisted of all the negative items. The factor structure proves that the negative items are not simply the inverse of positive items. That is, the negative item subscales are intended to reflect separate and different dimensions from those represented in the positive item subscales[1]
Responsiveness	
Discriminative (between individuals)	The CALTARS scales averaged over six sessions discriminate between individuals on pretherapy status when patient contributions scales are used. See Convergent validity section
Evaluative (within individual across time)	The scales were completed six times throughout the duration of therapy. Multivariate analysis of variance showed the mean patient therapist and judge ratings of patient and therapist positive contributions were higher in the 20th session than in the first and third sessions (Scheffé averaged F for patients = 5.6, $p < 0.001$; for therapists $F = 9.6$, $p < 0.001$)[1,2]
Acceptability	
Number of items	42
Administration method	Judge-completed rating scale; therapist and patient self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	French/German
Access by ethnic minorities	No details
Feasibility	
Copyright	1984, Williams and Wilkins Co.
Web or scanning options	No details
Training details	Two judges each with 6 years' clinical experience underwent 15 hours of training
Administration/process details	Patients and therapists completed the CALTARS scales immediately following sessions 1, 3, 5, 10, 15 and 20[1,2]
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Each of the items is rated on a five-point Likert-type scale of intensity for presence ranging from 0 (not present) to 5 (intensely present)
Normative data	No details
Résumé	
Strengths	The CALTARS scales permit views of the alliance from each of the perspectives of patient and therapist, as well as judge
	Therapist and patient ratings of the alliance predict outcome
Weaknesses	At 42 items the measure is lengthy; this could pose problems in implementing the measure in some service settings
Areas for further research	The CALTARS scales have been developed into the California Psychotherapy Alliance Scales of 24 items for patient, therapist and judge versions

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Primary references

1. Marmar CR, Horowitz MJ, Weiss DS, Marziali E. The development of the Therapeutic Alliance Rating System. In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford Press; 1986. pp. 367–90.
2. Marziali E. Three viewpoints on the therapeutic alliance. Similarities, differences, and associations with psychotherapy outcome. *J Nerv Ment Dis* 1984;**172**:417–23.
3. Bachelor A. Comparison and relationship to outcome of diverse dimensions of the helping alliance as seen by client and therapist. *Psychotherapy* 1991;**28**:534–49.

Secondary references

4. Bachelor A, Salame R. Participants' perceptions of dimensions of the therapeutic alliance over the course of therapy. *J Psychother Pract Res* 2000;**9**:39–53.
5. Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook*. New York: Guilford Press; 1986.
6. Hentschel U, Bijleveld CCJH. It takes two to do therapy: on differential aspects in the formation of therapeutic alliance. *Psychother Res* 1995;**5**:22–32.

C7
Capacity for Dynamic Process Scale (CDPS)

General details	
Author	Baumann BD
Language	English
Country of publication/development	USA
Publication date	2001
Publisher	No details
Purpose and overview	
The measure is intended to demonstrate a patient's ability to engage in psychodynamic psychotherapy, with respect to collaborating therapeutically regarding problems of an affective and interpersonal nature	
Theoretical orientation	Psychodynamic
Population details	Adults in short-term psychotherapy, outpatients[1,4]
Perspective	In addition to the therapist scores, an independent rater can score the interactions of therapist–client from video-recordings of the interviews
Measure used by	Clinicians
Other versions	No details
Notes	16 males, 22 females. Mean age 28.29 years. Mainly Caucasian (36/38)
	Mainly upper-lower class and lower-middle class
	18 single, 10 married, 10 divorced
	37/38 with DSM-IV axis I diagnosis, primarily mood disorder (22/38)[1]
Areas of therapist–patient interaction addressed: Map	
Framework: collaborative/participative/involving	
Emotional expression: expression of feelings	
Achieving a working relationship: working alliance; affective bond	
Changing view of self with others: corrective emotional experience	
Inferred from items listed in ref. 1	
Dimensions	
(1) Appears introspective; (2) integrates affect; (3) verbal fluency; (4) insight; (5) perceives affective aspects of problem; (6) differentiates affect; (7) differentiates personal events; (8) positive relationship; (9) collaborates therapeutically.	
Reliability	
The CDPS has adequate internal consistency. Inter-rater reliability across each item ranged from partial to adequate	
Split-half	No details
Internal consistency	Coefficient alpha of 0.83 was the value for the CDPS scale rated by the therapists (a value of 0.87 was presented by the external raters). For both therapist and external rater, all of the item-to-scale correlations substantially exceeded 0.30 and were significant at $p < 0.05$, thereby supporting the use of the CDPS as a unitary construct[1]
Inter-rater	ICC (one-way random effects model) of 0.89
	Inter-rater reliability was also calculated across each of the items, with scores ranging from 0.64 to 0.86[1]
Test–retest	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Validity	
Convergent validity ranged from partial to adequate for the Penn Therapist Facilitating Behavior's Questionnaire Method (Penn TFBq) with CDPS-therapist and CDPS-external rater, respectively In testing the discriminant validity, only two of 12 correlations were over 0.30, suggesting that the CDPS can demonstrate adequate discriminant validity	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	Correlations were calculated between the CDPS total score and two measures of therapeutic alliance (Penn TFBq, HAq-R) and with mixed results. Correlations were calculated between these scales and CDPS total score after the initial feedback session and after the third follow-up session. Only significant correlations were found between CDPS (both therapist and external rater) and Penn TFBq on initial feedback, with correlations of 0.46 ($p < 0.001$) and 0.53 ($p < 0.0001$), respectively[1] A five-variable stepwise regression model demonstrated a strong association between personality attributes assessed through the Rorschach and judgements of subjects' potential for engaging in dynamic psychotherapy[4]
Discriminant	A Pearson correlation was used to assess the relationship between the total CDPS score and a number of measures [Global Assessment of Functioning (GAF), Global Assessment of Relational Functioning (GARF) and the Social and Occupational Functioning Assessment Scale (SOFAS)]. All correlations were discriminative, except for the SOFAS with the CDPS-therapist correlation which was significant with a correlation value of 0.39 ($p < 0.01$) Self-report scales were used: SCL-90-R (GSI), Inventory of Interpersonal problems (IIP) and Social Adjustment Scale (SAS): a significant correlation was only found between the external rater-CDPS and the SAS at –0.38 ($p < 0.05$)[1]
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	9
Administration method	Interview
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

Feasibility	
Copyright	2001, Society for Psychotherapy Research
Web or scanning options	No details
Training details	No details
Administration/process details	The therapist produces nine individual scores and one overall score. These scores can then be compared with those of an external rater, who will view video-recordings of the therapist–client interviews[1]
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal. The measure assesses nine basic areas (see Dimensions), for which the clinician assigns a rating for each on a five-point likert scale (1 = minimal, 5 = maximal)
Normative data	No details
Notes	
Unable to obtain refs 2 and 3. Information from ref. 4 is based on that provided in the abstract only	
Résumé	
Strengths	Partially adequate to adequate convergent and discriminant validity, and inter-rater reliability. Adequate internal consistency
Weaknesses	More areas of reliability and validity must be addressed
Areas for further research	Further testing of psychometric properties

Primary references

1. Baumann BD, Hilsenroth MJ, Ackerman SJ, Baity MR, Smith CL, Smith SR, *et al*. The Capacity for Dynamic Process Scale: an examination of reliability, validity, and relation to therapeutic alliance. *Psychother Res* 2001;**11**:275–94.
2. Butler SF, Thackrey M, Strupp HH. Capacity for dynamic process scale: relation to patient variables, process and outcome in time-limited dynamic psychotherapy. Society for Psychotherapy Research, Ulm, Germany; June 1987.
3. Thackrey M, Butler SF, Strupp HH. Measurement of patients' capacity for dynamic process. Society for Psychotherapy Research, Evanston, IL; June 1985.

Secondary reference

4. Alpher VS, Perfetto GA, Henry WP, Strupp HH. The relationship between the Rorschach and assessment of the capacity to engage in short-term dynamic psychotherapy. *Psychotherapy* 1990;**27**:224–9.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C8
Carkhuff Scales

General details	
Author	Carkhuff RR
Language	English
Country of publication/development	USA
Publication date	1969
Publisher	Rinehart and Winston

Purpose and overview

A rating scale that includes the concept of empathy as well as other 'core' ingredients of helping.[6] The Truax and Carkhuff scales (1967) included the core conditions of empathy, warmth and genuineness. Later, Carkhuff (1969) added scales reflecting a more active therapy strategy, measuring aspects such as immediacy of relationship, facilitative self-disclosure and confrontation.[2] Developed for use in the client-centred framework[4]

The Carkhuff (1969) scales comprise the following subscales: empathic understanding, communication of respect, personally relevant concreteness, facilitative self-disclosure/genuineness, confrontation and immediacy of relationship

Theoretical orientation	Client-centred,[4] group psychotherapy,[18] Carkhuffian systematic human relations[6] and Rogerian relationship therapy[2]/pan-theoretical
Population details	See below
Perspective	Trained independent raters
Measure used by	Practitioners, researchers. Used as training tools[2]
Other versions	No details
Notes	*Practitioners:* Graduates,[2,19] psychotherapists,[4] high-functioning and low-functioning psychologists (Master's level),[6] counsellors,[8] undergraduates,[5,11,15] registered nurses who practice in an acute- and chronic-care hospital,[13] professionals[15] and novices[15] *Clients:* Late adolescents and adult outpatients (minimum 15 years) encountering difficulties such as vocational indecision, marital conflicts, social isolation and other interpersonal problems[2] College students in a counselling practicum course. 14 females. Average age 23.7. Almost all were functioning in some kind of residential or peer counselling capacity at the time[3] Cathy and Mike in Carl Rogers' videos[4] The role of the client was played by a trained male psychodramatist[5] Male and female applicants for counselling in a university counselling center[6] 31 undergraduates, ten male. Clients had a variety of emotional and personal problems deemed to require several interviews to resolve[8] Clients 8 and 12 years of age[14] Hospitalised mental patients[18] *Raters:* The raters were six judges, two each trained on counsellor empathy, respect and genuineness to a reliability of 0.80[4] Trained independent raters[5] Psychologists, psychiatrists and psychiatric social workers[18]

continued

Areas of therapist–patient interaction addressed: Map	
Therapist engagement: empathy/sensitivity; respect; openness	
Threats to the relationship: confrontations	
Inferred from subscale names	
Dimensions	
NA	
Reliability	
The inter-rater reliability of the Carkhuff scales is adequate. One study reports an improvement from partial adequacy to adequate inter-rater reliability from session 1 to session 2	
Split-half	NA
Internal consistency	NA
Inter-rater	Empathic understanding: 0.91 for the pair of raters trained by Carkhuff and 0.88 for the pair of raters trained by a doctoral-level psychologist[1]
	Immediacy and facilitative self-disclosure: Pearson product-moment correlation of 0.97[2]
	Session 1: peer Carkuff empathy 0.52, trained Carkhuff empathy 0.62 (0.84 Ebel intraclass correlation), trained Carkhuff gross 0.60 (0.86 Ebel intraclass correlation). Session 2: trained Carkuff gross 0.78 (0.89 Ebel intraclass correlation)[3]
	Two judges rating transcripts of Carl Rogers yielded reliability Ebel coefficients of 0.75 for counsellor empathy, 0.95 for counsellor respect and 0.86 for counsellor genuineness[4]
	Inter-rater reliabilities, computed across all subjects for each of the nine excerpts, ranged from 0.94 to 1.00 for the Carkhuff scale[5]
	Global level of facilitation scale (LOF): inter-rater reliability of 0.91[6]
	Empathic understanding: a minimum of 0.95 was maintained[9]
Test–retest	No details
Validity	
The concurrent validity of the Carkhuff scales with the Truax and Carkhuff scales is adequate	
The Carkhuff scales demonstrate partial predictive validity in predicting scores on the Tennessee total positive scale	
The construct validity of the Carkhuff scales is not adequate, as it cannot be measured readily with much agreement	
Inadequate to partial convergent validity has been demonstrated for the Carkhuff scales	
Face	Carkhuff's empathy scale is a truncated version of the Truax and Carkhuff scale[5]
Content	No details
Criterion (a) concurrent	Carkhuff Self Disclosure Scale correlated with Truax and Carkhuff Warmth Scale ($r = 0.70$) and their Genuineness Scale ($r = 0.87$), as well as with Carkhuff's Immediacy Scale ($r = 0.78$). Carkhuff Immediacy Scale correlated 0.78 with Carkhuff Self Disclosure Scale, 0.85 with Truax and Carkhuff Warmth Scale, and 0.87 with Truax and Carkhuff Genuineness Scale[2]
	The correlation between the combined sets of ratings for the Carkhuff scale and the Truax scale was 0.89 ($n = 42$, $p < 0.001$), indicating that the scales do account for much of the same variance. The correlations between the scales for each content-affect combination range from 0.45 to 0.73. Using a *z*-score transformation, no statistically significant differences were found on mean differences between scales[5]
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Criterion (b) predictive	The Carkhuff scales were assessed for their ability to predict therapy outcome, as measured by six outcome measures. Correlations were all positive, but only statistically significant between the scales and the Tennessee total positive score ($r = 0.42$, $p < 0.01$). Least squares multiple regression analyses were also conducted to determine whether the scales predicted outcome in combination with any of five other empathy measures. The scales and client-perceived empathy together produced a multiple correlation of 0.54 ($p = 0.02$) with changes on the Tennessee total positive score to account for 30% of this variance[8]
Construct	Empathic understanding: analysis of variance with one between-groups factor (empathy level) and one within-groups factor (trainer type) was applied to the ratings. Significant main effects were found for both trainer type ($F = 4.73$, $p < 0.05$) and empathy level ($F = 38.07$, $p < 0.05$). Post hoc t-tests that compared empathy scores between rater pairs indicated a significance difference only at high empathy levels ($t = 2.30$, $p < 0.05$). Rating differences were found only at higher empathy levels, with the Carkhuff-trained raters rating the therapist's responses as significantly less 'empathic' than the non-Carkhuff-trained raters. Several explanations are possible: (1) the results are a function of random individual differences; (2) scores tend to cluster at the extremes of the scale; (3) the construct of empathy as it is presently defined is not valid and therefore cannot be measured readily with much agreement. If the difficulty rests with the construct validity of the scale, serious questions arise about its utility as a research instrument[1] Ratings utilising more similar construct definitions (Carkhuff empathy and Carkhuff gross) agreed more than ratings based on dissimilar constructs (i.e. GAIT and Carkhuff)[3]
Convergent	Trained Carkhuff empathy significantly correlated with peer Carkhuff empathy (0.27, $p < 0.05$) and trained Carkhuff gross (0.47, $p < 0.001$) in session 1. Trained Carkhuff empathy as in session 1 did not significantly correlate with either GAIT rating or trained Carkhuff gross in session 2. In session 1, peer Carkhuff empathy correlated (0.29, $p < 0.05$) with trained Carkhuff gross, but there are no ratings from session 2. Session 1 trained Carkhuff gross correlated only with trained Carkhuff gross (0.24, $p < 0.05$) from session 2. The relationship between GAIT empathy and Carkhuff accurate empathy was negative and not significant (–0.10). It appears that at least some of the difference between GAIT and Carkhuff empathy ratings can be closed by employing (1) a construct which falls between the Carkhuff and GAIT empathies on the continuum ranging from specific to global, and (2) the same situational sample[3] Empathic understanding: the Carkhuff scale was significantly correlated with the Empathy Construct Rating Scale – 23 items[9]
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	Carkhuff scales demonstrated limited responsiveness. Empathic understanding: rating differences were found at only higher empathy levels, with the Carkhuff-trained raters rating the therapist's responses as significantly less 'empathic' than the non-Carkhuff trained raters. Several explanations are possible; (1) the results are a function of random individual differences; (2) scores tend to cluster at the extremes of the scale; (3) the construct of empathy as it is presently defined is not valid and therefore cannot be measured readily with much agreement[1] Observers rated high in LOF had a significantly ($F = 5.53$, df = 1,14, $p < 0.05$) higher percentage of agreement with client reports of their own feelings and concerns than those rated low in levels of facilitation (respective means 63.01, 58.57)[6]

continued

Evaluative (within individual across time)	Limited agreement between Carkhuff ratings from sessions 1 and 2. This may be because the trained Carkhuff raters may have recalled their ratings of session 1 while rating session 2[3]
Acceptability	
Number of items	NA
Administration method	Questionnaire/rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1969, Rinehart and Winston
Web or scanning options	No details
Training details	Raters should be well trained[5,8] Training on each scale took 10 hours and included conceptual discussion, rating of interview segments not used in the study itself, and meetings to discuss inter-judge discrepancies[4] One of the two judges was trained by Carkhuff and, in turn, trained the other judge[8] Untrained raters tended to agree with the more highly trained raters within each definition and session[3]
Administration/process details	Before testing, each subject was given a written explanation of the helping process, which included an explanation of the reflection of feeling response and its importance in the counselling interview. Subjects responded to nine simulated client statements, presented on videotape. The tape was stopped after each excerpt to allow subjects to respond. A copy of the videotape of simulated statements, and descriptions of the content and affect for each statement were provided for raters[5]
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert. The Empathic Understanding Scale is a five-point scale. Lower scores represent lesser degrees of empathic response
Normative data	No details
Notes	
The Carkhuff (1969) scales are a modification of the more comprehensive scale reported by Truax and Carkhuff (1967). Carkhuff (1969) shortened the original nine-point scales to five points to increase 'reliability' (p. 315) and clarified interchangeable, additive and subtractive counsellor responses to 'reduce ambiguity' (p. 315)[5] Previous experience shows that the Carkhuff scale can be used more reliably than the Truax and Carkhuff (1967) Accurate Empathy Scale[4]	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Résumé	
Strengths	Adequate inter-rater reliability and concurrent validity. Partially adequate predictive and convergent validity
Weaknesses	Few areas of reliability addressed. Inadequate construct validity. Limited responsiveness. Thorough training required
Areas for further research	Refine the rating scales for the facilitative conditions so that they are more firmly anchored to reference points of specific therapist behaviours[2]
	Future research should be directed towards assessing the interaction of interpersonal relationship skills and client statement qualities[5]

Primary references

1. Avery AW, Danish SJ. Assessing intertrainer effects on the empathic understanding scale: training the trainer. *J Clin Psychol* 1976;**32**:404–7.
2. Barrow JC. Interdependence of scales for the facilitative conditions: three types of correlational data. *J Consult Clin Psychol* 1977;**45**:654–9.
3. Dooley D, Lange AJ, Whiteley JM. Sources of discrepancy in Carkhuff and Gait measurements of empathy. *Psychother Theory Res Pract* 1979;**16**:337–44.
4. Edwards HP, Boulet DB, Mahrer AJ, Chagnon GJ, Mook B. Carl Rogers during initial interviews: a moderate and consistent therapist. *J Counsel Psychol* 1982;**29**:14–18.
5. Engram BE, Vandergoot D. Correlation between the Truax and Carkhuff scales for measurement of empathy. *J Counsel Psychol* 1978;**25**:349–51.
6. Genthner RW, Saccuzzo DP. Accuracy of perception of psychotherapeutic content as a function of observers' level of facilitation. *J Clin Psychol* 1977;**33**:517–19.
7. Helms JE. Development and interrelationship of the Truax and Carkhuff scales of interpersonal function. *Catalog of Selected Documents in Psychology* 1974;**4**:19–20.
8. Kurtz R, Grummond DL. Different approaches to the measurement of therapist empathy and their relationship to therapy outcomes. *J Consult Clin Psychol* 1972;**39**:106–15.
9. Layton JM, Wykle MH. A validity study of four empathy instruments. *Res Nurs Health* 1990;**13**:319–25.

Secondary references

10. Carkhuff RR. Helper communication as a function of helpee affect and content. *J Counsel Psychol* 1969;**16**:126–31.
11. Dalton RF, Sundblad LM, Hylbert KW. An application of principles of social learning to training in communication of empathy. *J Counsel Psychol* 1973;**20**:378–83.
12. Hountras PT, Anderson DL. Counselor conditions for self-exploration of college students. *Personnel Guid J* 1969;**48**:45–8.
13. LaMonica EL, Carew DK, Winder AE, Haase AM, Blanchard KH. Empathy training as the major thrust of a staff development program. *Nurs Res* 1976;**25**:447–51.
14. Mook B. Analyses of therapist variables in a series of psychotherapy sessions with two child clients. *J Clin Psychol* 1982;**38**:63–76.
15. Pope B, Nudler S, Vonkorff MR, McGhee JP. The experienced professional interviewer versus the complete novice. *J Consult Clin Psychol* 1974;**42**:680–90.
16. Richardson BK, Smith M, Bolton B. Development of example-anchored scales of interpersonal functioning to assess rehabilitation counseling. *Rehabil Counsel Bull* 1974;**17**:188–97.
17. Tosi DJ, Eshbaugh DM. A cognitive-experiential approach to the interpersonal and intrapersonal development of counselors and therapists. *J Clin Psychol* 1978;**34**:494–500.
18. Truax CB. The process of group psychotherapy: relationship between hypothesized therapeutic conditions and intrapersonal exploration. *Psychol Monogr* 1961;**75**:35.
19. Walker RB, Latham WL. Relationship of a group counseling course, hours in counselor education, and sex to empathic understanding of counselor trainees. *Counsel Educ Supervis* 1977;**16**:269–74.

C9
Child Psychotherapy Process Scales (CPPS)

General details	
Author	Estrada AU
Language	English
Country of publication/ development	USA
Publication date	1996
Publisher	NA
Purpose and overview	
Developed from the Vanderbilt Psychotherapy Process Scale (VPPS; Suh, Strupp and O'Malley, 1986), the CPPS assesses child therapy process. It was designed to assess both positive and negative aspects of child and therapist behaviours and attitudes as displayed within whole sessions or segments thereof. These attitudes and behaviours are likely to facilitate or impede progress in child therapy. This measure is relevant to both therapists and children, with the 33 items being split between those relevant to child attitudes and behaviours, and 18 to those of the therapist	
Theoretical orientation	Psychodynamic: the predominant theoretical orientation of the supervisors of the training centre was psychodynamic. A combination of verbal and play therapy was used; long-term individual therapy[1]
Population details	Clinical children. See below
Perspective	Independent rater: the measure was designed to be used by clinical psychology graduate students with minimal clinical experience (objective observers)
Measure used by	Practitioners (mainly psychologists) and for training purposes
Other versions	None, but was originally called the Loyola Child Psychotherapy Process Scales
Notes	*Practitioners:* In development study: 13 graduate student therapists (nine psychology, two social work, two pastoral studies) (9 F, 4 M)[1] *Clients:* Clinical children. Characteristics of sample: gender: 9M, 4F; age: 6–12 years; treatment history: been in therapy for 4–18 months; diagnosis: included oppositional defiant disorder, ADHD and PTSD[1] *Raters:* Psychologist: two clinical psychology graduate students (and the author) acted as raters of the emerging recorded transcripts. 35 sessions were evaluated and 105 segments of therapy were evaluated from the transcripts[1]
Areas of therapist–patient interaction addressed: Map	
Therapist engagement	
Patient engagement	
Therapeutic techniques	
Achieving a working relationship	
Dimensions	
Child items:	
Child therapeutic relationship	Eight items
Child therapeutic work	Five items
Child therapeutic readiness	Four items
Therapist items:	
Therapist technical work	Ten items
Therapist therapeutic relationship	Five items
Therapist technical lapse	Three items

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
Internal consistency values for the CPPS ranged from inadequate to adequate, the majority of results being adequate	
The CCPS demonstrates adequate inter-rater reliability	
Split-half	No details
Internal consistency	Cronbach's alpha – for child scale: child therapeutic relationship 0.83, child therapeutic work 0.82, child therapeutic readiness 0.68. Internal reliabilities for therapist scale: therapist technical work 0.88, therapist therapeutic relationship 0.73, therapist technical lapse 0.49[1]
Inter-rater	Correlation = 0.77. Agreement between raters averaged at 0.77 (0.59 < *r* <0.94) for both the child and therapist items
Test–retest	No details
Validity	
The results of a factor analysis of the CPPS suggest a three-factor solution to both the child and therapist subscales	
Face	11 items were adapted from the VPPS, and experienced clinical child psychologists generated additional items
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	Principal components analysis with oblimin (oblique) rotation was applied to both the 15-item child scale, and the 18-item therapist scale. Kaiser-Meyer-Olkin (KMO) Statistic for child (0.86), therapist (0.90). Using eigenvalues >1.0: – Child: three-factor solution accounting for 73.4% of the variance. 1st factor (child therapeutic relationship) accounted for 47.4% of variance, with 8/15 items loading >0.48. 2nd factor (child therapeutic work) 18.2% variance, 5/15 items loading >0.67. 3rd factor (child therapeutic readiness) 7.8% variance, 3/15 items loading >0.49. One cross-loading between factor 1 (–0.50) and factor 3 (–0.58) – Therapist: three-factor solution accounting for 66% total variance. First factor (therapist technical work) 40.3% variance, with 10/18 items loading, nine >0.63, one at 0.35. Second factor (therapist therapeutic relationship) 18.5% variance, 5/18 items loading >0.57. Third factor (therapist technical lapse) 7.2% variance, 3/18 items loading >0.63. No cross-loadings. Intercorrelations between child dimensions: 0.27–0.40; therapist dimensions 0.19–0.34. Intercorrelations between child and therapist dimensions: 0.08–0.88[1]
Responsiveness	
Discriminative (between individuals)	High-quality sessions (as identified by nine segments comprising the three sessions with the most favourable CPPS ratings) were compared with low-quality sessions (as identified by nine segments comprising the three sessions with the least favourable CPPS ratings).[2,3] Therapist and child client utterances from the high- and low-quality sessions were rated (using different raters from ref. 1) on 15 language interaction scales derived from the Stuttgart International Category System (SICS) (Czogalik *et al*., 1987). For the therapist discourse data,[2] confirmatory factor analysis revealed large statistically reliable differences between high and low quality sessions (nearly 3 SD). For the client discourse data,[3] the difference in high- and low-quality sessions was even greater, supporting the discriminate validity of the CPPS
Evaluative (within individual across time)	No details

continued

Acceptability	
Number of items	33
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	No details
Web or scanning options	No details
Training details	No details
Administration/process details	No details
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert. Five-point scales indicating the extent to which the characteristic is present
Normative data	No details
Résumé	
Strengths	Results indicate that the child and therapist CPPS factors are reliable and can discriminate child therapy processes from sessions differing in judged quality.[1–3] The CPPS factors describe clinically significant and valid constituents of child psychotherapy[1]
Limitations	Some of the factors (e.g. child therapeutic readiness) will require generation of more items to increase reliability and consistency. Findings are constrained by the small sample size, narrow band of disorders, use of transcripts only, and the fact that therapy was conducted by trainees supervised from a psychodynamic model[1]
Future research	Future efforts should attempt to cross-validate the findings with larger, diverse samples of clients and therapists.[1] More psychometric properties need to be addressed

Primary reference

1. Estrada AU, Russell RL. The development of the Child Psychotherapy Process Scales (CPPS). *Psychother Res* 1999;**9**:154–66.

Secondary references

2. Russell RL, Byrant FB, Estrada AU. Confirmatory P-technique analyses of therapist disclosure: high- versus low-quality child therapy sessions. *J Consult Clin Psychol* 1996;**64**:1366–76.

Additional reference may now be published:

3. Russell RL, Estrada AU, Byrant FB. Updating P-technique analyses of child therapy processes: bootstrapping, confirmatory and discriminant analyses. Manuscript submitted for publication; 1998.

Potentially useful reference about a different scale measuring relevant therapist patient interaction concepts, rated from child and therapist's perspective (summary in ref. 1):

4. Smith-Acuna S, Durlak J, Kaspar C. Development of child psychotherapy process measures. *J Clin Child Psychother* 1991;**20**:126–31.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C10
Client Attachment to Therapist Scale (CATS)

General details	
Author	Mallinckrodt B
Language	English
Country of development	Oregon, USA
Publication date	1995
Publisher	NA
Purpose and overview	
The Client Attachment to Therapist Scale views the therapeutic relationship from an attachment perspective, which measures the quality of client's attachment to their therapist, in terms of feelings and attitudes	
Theoretical orientation	Attachment theory/psychoanalytic/counselling psychology
Population details	Adults
Perspective	Client rated
Measure used by	Psychotherapists
Other versions	No details
Notes	Details on participants involved in the piloting/development of the questionnaire:[1] Clients were solicited for participation during a 3-year period from four counselling agencies including a university counselling centre, a community college counselling centre, a hospital-based outpatient clinic and an in-house training clinic operated by a counselling psychology programme. Therapists were senior staff interns or graduate students in training at these agencies
Areas of therapist–patient interaction addressed: Map	
Therapy context: boundaries; influence; power/coercion; responsibilities	
Roles: friend/companion; attachment figure; confidant; good object; protector	
Individual differences: attachment styles; overprotective; defensive style/repression	
Therapist engagement: empathy/sensitivity; warmth; genuineness; respect; support/tolerance; listening; hope/encouragement	
Patient engagement: expectation/preferences; attraction	
Framework: reciprocal; collaborative/participative/involving; controlling	
Therapeutic techniques: transference; responsiveness/receptiveness/attunement; ruptures/repair	
Threats to the relationship: intrusive; defensive; critical; hostility/anger; fear; sexual involvement; hidden agendas; resistance; confrontations; withdrawal	
Information derived from items	
Dimensions	
Secure	14 items: experiencing the therapist as responsive, sensitive and understanding and emotionally available; feeling hopeful and comforted by the counsellor; and feeling encouraged to explore frightening or troubling events
Avoidant–fearful	12 items: suspicion that the therapist is disapproving, dishonest and likely to be rejecting if displeased; reluctance to make personal disclosures in therapy; and feeling threatened, shameful and humiliated in the sessions
Preoccupied–merger	10 items: longing for more contact and to be 'at one' with the therapist, wishing to expand the relationship beyond therapy, and preoccupation with the therapist and the therapist's other clients

continued

Reliability	
Cronbach's alpha was used as an estimate of internal consistency. There was partial support for the measure's internal reliability: one of the scales demonstrated adequate consistency (preoccupied–merger), while the other two demonstrated partial adequacy One of the scales demonstrated adequate internal consistency and the other two partial internal consistency All of the scales demonstrated adequate test–retest reliability at 3.24 weeks	
Split-half	No details
Internal consistency	Internal consistencies for the scales were as follows: secure 0.64, avoidant–fearful 0.63, preoccupied–merger 0.81
Inter-rater	NA
Test–retest	The second CATS survey was completed by 17 clients a mean of 3.24 weeks after the first. Test–retest correlations were: secure 0.84, avoidant–fearful 0.72, merger–merger 0.86.
Validity	
Procedures to ensure content and face validity were carried out Evidence of concurrent validity was reported with the Working Alliance Inventory (WAI), the Bell Object Relations and Reality Testing Inventory (BORRTI) and the Self Efficacy Scales (SES). See ref. 1 for full details. To summarise: • Adequate concurrent validity with the WAI scale was demonstrated by the secure and avoidant–fearful scales • The avoidant–fearful scale demonstrated partially adequate concurrent validity with the alienation, egocentricity and social incompetence scales of the BORRTI • The avoidant–fearful and preoccupied–merger scales demonstrated partially adequate concurrent validity with the SES scales • None of the CATS scales displayed adequate/partially adequate concurrent validity with the Adult Attachment Scales (AAS)	
Face	Nine experienced therapists generated items for CATS, redundant items were removed and some new items generated. Unclear items were reworded by graduate students to ensure face validity
Content	In the process of reducing the pool of 75 items to 36, items were removed from the subscales if (1) the item seemed conceptually unrelated to other items on the subscale, or (2) the item was highly correlated or seemed redundant with another item on the subscale
Criterion (a) concurrent	CATS scales correlated with WAI scales at a range –0.56 to 0.82[1] CATS scales correlated with BORRTI scales at a range –0.29 to 0.46 CATS scales correlated with AAS scales at a range of –0.10 to 0.18 CATS scales correlated with SES scales at a range of –0.39 to 0.15
Criterion (b) predictive	No details
Convergent	See under Concurrent
Discriminant	No details
Factor structure	A principal factors analysis was used. The resulting secure, avoidant–fearful and preoccupied–merger factors accounted for 26%, 7% and 5% of variance in the data, respectively The avoidant–fearful and secure subscales were significantly negatively correlated ($r = -0.51$, $p < 0.01$) and the secure and preoccupied–merger scales were positively correlated ($r = 0.23$, $p < 0.01$). The avoidant–fearful and preoccupied–merger scales were not significantly correlated ($r = -0.10$).

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Responsiveness	
Discriminative (between individuals)	Clients varied considerably with regard to the number of sessions they had completed at data collection, thus the sample was divided into three groups of approximately equal size based on length of therapy. Comparisons suggest that secure subscale scores are significantly different depending on the length of therapy, ($F_{2,135} = 4.26$, $p < 0.05$). Duncan's multiple range test, used for follow-up group comparisons, indicated that clients seen for five to eight sessions at the time of data collection had significantly lower secure subscale scores than either of the other two groups seen for a longer period
Evaluative (within individual across time)	No data over time were collected
Acceptability	
Number of Items	36
Administration method	Self-report questionnaire
Time taken to complete	Not specified
Flesch reading age	Not specified
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1995, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	The CATS is prefaced with these instructions: "These statements refer to how you currently feel about your counsellor. Please try to respond to every item using the scale below to indicate how much you agree or disagree with each statement"
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Likert scale ranging from 1 = strongly disagree to 6 = strongly agree. 25 items were changed to the negative to minimise response-set bias
Normative data	See ref. 1

Notes

Limitations of study: all measures were self-report; factor analysis would be more reliable with a greater sample size, generalisabilty of findings is poor and CATS was completed anonymously, which restricts the usefulness of the CATS for clinical purposes in ongoing therapy[1]

CATS measures constructs that are different in important ways to the working alliance.[2] Alliance measures of the bond between client and therapist typically measure only the bond's relative strength, whereas the CATS taps several possible underlying dimensions of weak bonds

The relatively low internal consistencies of the two subscales suggest that there may be subsets of different strategies within an overall theme of secure or avoidant–fearful attachment

The authors agree with Robbins, that the preoccupied–merger and avoidant–fearful subscales may capture positive and negative aspects of transference

Highlights the low internal consistency of the first two factors, secure and avoidant–fearful (alpha coefficients of 0.63 and 0.64), which raises questions about the unidimensionality of the scale and their exact interrelationship.[3] One possibility is that the items reflect a composite of positive and negative reference points to both self and others. Griffin and Bartholomew (1994) argued that working models of self and others must be separated and understood as independent bipolar dimensions. Mallinckrodt *et al*. (1995) need to explore further the unidimensionality, stability and the meaning of the CATS within this context

continued

Résumé	
Strengths	Differs from other alliance measures in that the CATS measures the client–therapist relationship from an attachment perspective
	The secure and avoidant–fearful scales display convergent validity with the WAI, but the low amounts of variance shared with the WAI by the preoccupied–merger scale show that the measure is measuring an aspect of the counselling relationship that is distinct from the working alliance
Weaknesses	Length of measure (36 items) could make it difficult to use in routine practice over time
	The internal consistencies of the secure and avoidant–fearful scales are only partially adequate
Weaknesses	CATS was completed anonymously in a pilot study,[1] which means that the usefulness of CATS for clinical purposes in ongoing therapy is restricted
	Low numbers of males and ethnically diverse clients restricts generalisability of findings
Areas for further research	Further testing of psychometric properties
	CATS needs to be adapted for clinical purposes in ongoing therapy

Primary references

1. Mallinckrodt B, Gantt DL, Coble HM. Attachment patterns in the psychotherapy relationship: development of the client attachment to therapist scale. *J Counsel Psychol* 1995;**42**:307–17.
2. Mallinckrodt B, Coble HM, Gantt DL. Toward differentiating client attachment from working alliance and transference: reply to Robbins (1995). *J Counsel Psychol* 1995;**42**:320–2.
3. Robbins SB. Attachment perspectives on the counselling relationship: comment on Mallinckrodt, Gantt, and Coble (1995). *J Counsel Psychol* 1995;**42**:318–19.

Secondary references

None

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C11
Client Resistance Scale (CRS)

General details	
Author	Mahalik J
Language	English
Country of publication/development	USA
Publication date	1994
Publisher	NA

Purpose and overview	
The measure was designed to identify the salient dimensions of resistance that manifest themselves in client dialogue. The instrument looks at whether the client opposes expression of painful affect, opposes recollection of material, opposes the therapist, opposes change and opposes insight in therapy	
Theoretical orientation	Analytical psychotherapy, cognitive behavioural, gestalt, rational emotive therapy
Population details	Adult
Perspective	Judge rated
Measure used by	Research therapists/clinicians
Notes	The measure was developed using the set of films *Three approaches to psychotherapy*, which show Carl Rogers, Frederick Perls and Albert Ellis each conducting individual therapy with a female client named Gloria, and the set of films *Three approaches to psychotherapy III*, which show Donald Meichenbaum, Aaron Beck and Hans Strupp each conducting individual therapy with a male client named Richard. 16 master's level graduates in counselling psychology from a private eastern university (15 women and one man) served as raters on this study

Areas of therapist–patient interaction addressed: Map

Therapy context: type of therapy; boundaries

Individual differences: defensive style/repression

Patient engagement: motivation; commitment; intentions

Framework: congruent

Therapeutic techniques: exploration; ruptures/repair

Threats to the relationship: defensive; resistance; withdrawal

Information derived from description of dimensions

Dimensions	
Opposing expression of painful affect	Examines the extent to which the client blocks the expression of painful feelings
Opposing recollection of material	Designed to examine the extent to which the client provides vague versus detailed information to the therapist about himself or herself; tendency to avoid self-disclosing communication
Opposing therapist	Designed to examine the extent to which the client complies with the therapist in pursuing the therapeutic task as set by the therapist
Opposing change	Reflects the client's expressed desire to change and the extent of the client's satisfaction with the status quo of his/her life circumstances
Opposing insight	Reflects the extent of the client's self-understanding in terms of making connections between his or her experiences, thoughts, feelings and behaviours

continued

Reliability	
r_k was calculated to examine the reliability of the judges using the subscales of the CRS in rating client speaking turns (CST) and to assess the reliability of different numbers of judges using the CRS. See ref. 1 for formula	
The subscales demonstrated adequate inter-rater reliability in rating CST from all sessions combined for four judges, three judges and two judges. Reliability ratings of the subscales for single judges from all sessions combined ranged from adequate to partially adequate	
Reliability of the mean ratings for CST for individual sessions showed adequate reliability when scores for four and three judges were examined. Reliability of the mean for two judges or for single judges was still adequate for several subscales but less reliable. This was most evident for the 'Opposing insight' subscale. Therefore, it is recommended that at least three judges be used to obtain adequate reliability when rating CST with the CRS for individual sessions	
Split-half	NA
Internal consistency	NA
Inter-rater	*Over all sessions* Four judges: range 0.83 to 0.96 Three judges: range 0.79 to 0.95 Two judges: range 0.71 to 0.92 One judge: range 0.55 to 0.86 *Single sessions* Four judges: range 0.69 to 0.94 Three judges: range 0.62 to 0.96 Two judges: range 0.52 to 0.95 One judge: range 0.36 to 0.90
Test–retest	No details
Validity	
The CRS demonstrates adequate concurrent validity with the Hill Counselor Response Modes Verbal Category System (HCRMVCS; Hill, 1985). The construct validity of the CRS was supported by the finding that the measure discriminated between clients, therapists and therapist response modes	
Content validity issues were addressed by using a panel of judges to use the scales to rate tapes unrelated to the study	
Face	No details
Content	After the initial five scales were developed judges rated therapy tapes using the scales. Through feedback and discussion with judges, inconsistencies with the descriptors of the subscales were identified
Criterion (a) concurrent	Differences in resistance scores were found between the therapist response modes of the HCRMVCS
Criterion (b) predictive	No details
Construct	The subscales were differentially affected ($p < 0.05$) by the client, the therapist and therapist response mode, which gives some support for the construct validity of the CRS (see ref. 1 for details)
Convergent	See under Concurrent
Discriminant	No details
Factor structure	Pearson product-moment intercorrelations of the five subscales were low to moderate, ranging from 0.31 to 0.62, and all were significant at the 0.01 level, two tailed. This result suggests that the CRS is unidimensional. However, subscales were differentially affected by therapist, client, and therapist response mode, showing that examining the subscales is important
Responsiveness	
Discriminative (between individuals)	Significant differences were found in resistance scores for client, therapist and therapist response mode. These results suggest that the CRS can be used to descriminate between clients, therapists and therapist response mode (response modes were taken from the HCRMVCS)
Evaluative (within individual across time)	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Acceptability	
Number of Items	Five subscales rated by judges
Administration method	Trained judges rate client speech
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1994, American Psychological Association
Web or scanning options	No details
Training details	Judges received more than 20 hours' training in the use of the CRS, which included receiving didactic instruction, viewing videotapes not used in the study, and discussing and receiving feedback about the ratings
Administration/process details	Judges rate client speech, which varies in length from CST to whole sessions on each of the five subscales using audiotape, videotape, typed transcript or a combination of these
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Each of the five subscales is rated on a seven-point scale, with higher ratings reflective of greater amounts of the subscale being rated
Normative data	No details
Résumé	
Strengths	Adequate inter-rater reliability with three or more judges across combined sessions and for single sessions
	Concurrent validity with the HCRMVCS
Weaknesses	CRS has only been validated with videotaped films featuring other therapists and two clients, so reliability and validity information is difficult to generalise
	Use of the CRS requires at least three judges, each requiring 20 hours' training
	No research on application/use
Areas for further research	Further examination of validity evidence with actual clients
	Research examining change over time in client resistance levels
Primary reference	
1. Mahalik JR. Development of the Client Resistance Scale. *J Counsel Psychol* 1994;**41**:58–68.	
Secondary references	
None	

C12
Coding the Interaction in Psychotherapy (CIP)

General details	
Authors	Schindler L, Hohenberger-Sieber E, Hahlweg K
Language	English
Country of publication/development	Munich, West Germany
Publication date	1989
Publisher	NA
Purpose and overview	
The CIP was designed to code the interaction behaviours in psychotherapy. The CIP aims to analyse the moment-to-moment interactions to determine specifically the therapist (verbal) behaviours that affect clients and how these skills interact with behaviour change techniques	
Theoretical orientation	Behaviour therapy
Population details	Adults
Perspective	Rater
Measure used by	Therapists/research clinicians
Other versions	No details
Notes	The CIP was developed on an analysis of intake sessions
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: empathy/sensitivity; support; listening; hope/encouragement; praise/affirmation	
Patient engagement: motivation; commitment; intentions	
Maintaining the relationship: convergent; complementary; reciprocal; collaborative; congruent; structuring; directive	
Therapeutic techniques: responsiveness/receptiveness/attunement; exploration; feedback	
Non-verbal communication: silence	
Threats to the relationship: critical; hostility; resistance; confrontations; withdrawal	
Outcomes: working alliance; emotional expression; changing view of self with others	
Information derived from description of verbal behaviours	
Verbal behaviours	
Therapist empathy	Addressing emotions; reformulation; understanding
Therapist support	Confidence giving; positive feedback; minimal support
Therapist exploration	Information seeking; summarising
Therapist explanation	Neutral statement; structuring; explanation; self-disclosure
Therapist directivity	Directive guidance; instruction/advice
Therapist classification	Confrontation; interpretation; criticism
Client self-disclosure	Expression of negative feelings; expression of positive feelings
Client problem description	Background information; problem description
Client short answers	Short answers
Client change reports	Attempted self-control; reports in success; insight
Client cooperation	Goal formulation; expression of confidence; proposals for change
Client information seeking	Addressing therapeutic relationship; request for information
Client resistant behaviour	Avoidance/refusal; criticism/provocation; resignation

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
Adequate inter-rater reliability was proven, with strong correlations between judge and expert ratings	
Split-half	No details
Internal consistency	No details
Inter-rater	Inter-rater reliability levels were obtained by comparing each judge's rating with an 'expert' rating, obtained by consensus of three clinical psychologists involved in developing the CIP. Reliabilities were higher for categories with higher frequencies. The mean kappa coefficients for all raters and transcripts were 0.80 (SD 0.06) and 0.79 (SD 0.11) for therapist and client categories, respectively
Test–retest	No details
Validity	
Convergent validity with therapist and client ratings ranged from inadequate to partial, as demonstrated by low but significant correlations	
Adequate content validity was demonstrated by the CIP showing a shift in therapist and client activities over the course of the interview. The difference in the relative frequencies of categories was compared and found to be statistically significant, although the exact significance levels are not reported	
Face	No details
Content	The intake sessions were standardised, beginning with an explorative style and ending in more explanation. To test for content validity of the CIP, the frequencies of categories in the first and second halves of the interviews were compared using the non-parametric Wilcoxon test. The results indicated that the CIP did reflect the shift in therapist activity: therapists had significantly higher rates of explorative categories in the first half and explanation categories in the second half. Accordingly, clients had significantly higher rates of problem description in the first half and goal formulation in the second half
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	The CIP categories were correlated with corresponding client and therapist subjective ratings of each other. Seven of the 18 client CIP behaviours significantly correlated with therapist perceptions of the client and five of the 19 therapist CIP behaviours had significant correlations with client perceptions of the therapist (range 0.21 to 0.37)
Discriminant	No details
Factor structure	NA
Responsiveness	
Discriminative (between individuals)	The Mann–Whitney *U* test compared more and less experienced therapists. Significant differences emerged for six therapist categories and two client categories
Evaluative (within individual across time)	CIP used in process studies[2] to evaluate the influence of in-session behaviour on treatment outcome
Acceptability	
Number of items	37 verbal behaviours (19 therapist, 18 client)
Administration method	Rating scale
Time taken to complete	An average of 4 hours to code one tape
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

Feasibility	
Copyright	1989, The British Psychological Society
Web or scanning options	No details
Training details	Raters were three postgraduate psychology students. They were trained for 50 hours on coding. Periodic training sessions were held to prevent observer drift. A manual is available from the authors
Administration/process details	Trained raters code taped sessions. The sessions are rated for the frequencies of therapist and client behaviours
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	The scoring unit is a verbal response, which corresponds to the definition of a particular category. Units are compared between categories
Normative data	No details
Notes	
The validation of the CIP was conducted using intake sessions. The CIP is tailored to the therapy process, meaning that in this study not all of the categories would be relevant. There is potential for the CIP to be used in process studies to analyse the influence of in-session behaviour on treatment outcome	
Résumé	
Strengths	Adequate inter-rater reliability
Weaknesses	Burden on users: takes 4 hours to code one tape and 50 hours to train coders
Areas for further research	Further testing of psychometric properties
Primary reference	
1. Schindler L, Hohenberger-Sieber E, Hahlweg K. Observing client–therapist interaction in behaviour therapy: development and first application of an observational system. *Br J Clin Psychol* 1989;**28**:213–26.	
Secondary reference	
2. Schindler L. Social Influence and therapeutic change: interaction patterns of client and therapist correlated with treatment outcome. In Emmelkemp P, Florin I, Marks I, editors. *Theory and practice in behaviour therapy*. Lisse: Swets and Zeitlinger; 1988.	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C13
Coherence of the Relationship Theme (CRT)

General details	
Author	Mitchell J
Language	English
Country of publication/development	USA
Publication date	1995
Publisher	No details
Purpose and overview	
To observe empirically the tendency of different patients to link their Core Conflictual Relationship Theme (CCRT) components	
The additional scoring step allows a determination of a patient's capacity to form and describe links or interactions	
Theoretical orientation	Psychotherapy
Population details	See below, and case studies provided
Perspective	Independent rater
Measure used by	Psychiatric inpatients; outpatients
Other versions	No details
Notes	*Clients:* Psychiatric inpatients: suffered from severe psychopathology; all but one received a consensual diagnosis of schizophrenia or affective disorder from a primary therapist trainee and an attending psychiatrist *Outpatients:* no history of psychiatric hospitalisation, received consensual diagnoses of character disorders from a research assistant and the treating psychotherapist. These patients were treated by weekly psychotherapy for a 40-week period as part of a study of short-term therapy techniques[1]
Areas of therapist–patient interaction addressed: Map	
In this context, 'relatedness' is defined as a patient's willingness to affect others, and to be affected by them, as revealed in the patient's descriptions of his or her own encounters	
No other details available	
Dimensions	
Wishes	No further details
Responses from the other	
Responses from the self	
Reliability	
There is partially adequate inter-rater reliability between the two judges. No other areas of reliability were addressed	
Split-half	NA
Internal consistency	NA
Inter-rater	CRT link percentage (Lk%), which is the percentage of CCRT elements that were found to be linked, was judged by two raters, with an inter-rater reliability of 0.67
Test–retest	No details

continued

Validity	
The CRT correlates significantly with the diagnostic group variable and with levels of integrative failure. The CRT is able to discriminate between inpatient and outpatient groups	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	The diagnostic group variable, which ranks patients having an affective illness above patients having schizophrenia, correlates significantly with CRT Lk%, as does the levels of integrative failure (LIF; Grand *et al*., 1993) measure of differentiation
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	Patients in the outpatient group make a significantly greater percentage of links than patients in the inpatient group ($t = 3.134$, $p < 0.05$). Patients who linked only about 30% of their CCRT elements typically carried a diagnosis of schizophrenia, with its minimal capacity for relatedness. Where the CRT Lk% was much higher (approaching 60%), patients functioned fairly effectively in independent lives
Evaluative (within individual across time)	No details
Acceptability	
Number of items	NA
Administration method	No details
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1995, Lawrence Erlbaum Associates
Web or scanning options	No details
Training details	No details
Administration/process details	Instead of selecting relationship episodes (REs) for analysis, the scorer is instructed to identify all wishes and responses using the usual CCRT approach, whether or not they fall into what would have been proper REs. Then, the scorer performs the new additional step of determining whether or not each component is linked with another, e.g. a wish is considered to be linked if it has a person other than the patient as its object, and unlinked if the patient is the object
Support from measure developers	A copy of CRT scoring instructions, with many examples, is available from the author
FAQ facility	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Precision	
Scale type	NA
Normative data	No details
Résumé	
Strengths	Partially adequate inter-rater reliability. The CRT demonstrates the ability to distinguish significantly between inpatient and outpatient groups This pilot study offers support for the proposal that the frequency of linkage among CCRT elements, the CRT, is a meaningful measure of a capacity for relatedness and a useful enhancement of the CCRT system[1]
Weaknesses	More areas of reliability and validity need to be addressed
Areas for further research	See above
Primary reference	
1. Mitchell J. Coherence of the relationship theme: an extension of Luborsky's core conflictual relationship theme method. *Psychoanaly Psychol* 1995;**12**:495–512.	
Secondary references	
None	

C14
Core Conflictual Relationship Theme (CCRT)

General details	
Author	Luborsky L
Language	English
Country of publication/development	USA
Publication date	1976
Publisher	NA
Purpose and overview	
The CCRT is an attempt to capture some essential components of the individual's belief structure regarding others: what the individual frequently wishes from others, how he or she construes others as responding to such wishes, and how he or she responds to the interaction as construed. CCRT method is the assessment of which components occur with the greatest frequency across various relationships	
Theoretical orientation	Psychoanalytic/psychodynamic
Population details	Adults
Perspective	Judge
Measure used by	Clinicians/research clinicians
Other versions	*Additional CCRT procedures* Relationship Anecdote Paradigms Interview (RAP) CCRT Self-Report Questionnaire Self-Interpretation of the CCRT A CCRT scoring method based on the sequence of components
Notes	
Areas of therapist–patient interaction addressed: Map	
Individual differences: problem complexity; attachment styles; defensive style/repression	
Patient engagement: motivation; attraction; commitment; intentions	
Therapeutic techniques: transference	
Outcomes: emotional expression; changing view of self with others	
Changing view of self with others	
Information is derived from standard categories	
Dimensions	
Wishes (W)	
Responses of the main other person (RO)	
Responses of the self (RS)	
Reliability	
The inter-rater reliability of the CCRT across studies ranges from partial to adequate	
Split-half	NA
Internal consistency	NA
Inter-rater	Independent judgements of completeness of relationship episodes were satisfactory. The pooled judge intraclass correlation for 111 episodes was 0.68 ($p < 0.01$). The agreement was also satisfactory on the selection by two independent judges of the main other person with whom the self was interacting. For 80 episodes, 89% had the same other person identified by both judges

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	In the location of REs, the judges differed by an average of only 4.8 lines at the beginning and 7.9 lines at the end[6,27] Reliabilities for standard categories as measured by weighted kappas Wish and negative response of self: 0.61; negative response from other: 0.70[6] *Mean weighted kappa for seven samples* Wishes: 0.63 Responses from others: 0.66 Responses of self: 0.69[27] *Mean weighted Kappa for eight samples* Wishes: 0.61 Responses from others: 0.67 Responses from self: 0.71[27,28]
Test–retest	NA
Validity	
Good concurrent validity with ten of Freud's observations on transference has been demonstrated The CCRT has demonstrated partial predictive validity with improvement in the alliance Convergent validity with other formulation methods has been demonstrated	
Face	CCRT method was developed by Luborsky's observing and tracking of how he did the job of inferring the general relationship pattern from sessions, and monitoring how he inferred the general relationship pattern
Content	See above
Criterion (a) concurrent	The correspondence of Freud's observations on transference with CCRT evidence for 17 of Freud's 22 observations has been examined. For ten of Freud's observations the authors found good correspondence with CCRT evidence. For six, promising correspondence was found and for one, mixed evidence was found. For further details see ref. 27
Criterion (b) predictive	Accurate interpretations based on convergence of the interpretation with the independently established CCRT's wish and response from other were associated with improvement in the alliance during treatment Pearson correlations were as follows: rated benefits: 0.38, ($p < 0.05$); residual gain 0.44, ($p < 0.01$)[15,37]
Construct	There is support for the hypothesis that the CCRT method is related to Freud's transference template (see Convergent validity)
Convergent	*Similarity of CCRT with other formulation methods* Two judges rated the degree to which each one of the seven formulation methods was similar to the others. A 1–7 scale was used, where 1 means completely dissimilar, 7 is completely similar and 4 is somewhat similar. The two clinical judges agreed well with each other; their similarity ratings were correlated 0.74 ($p < 0.001$) The mean similarity rating of the CCRT was 4.92, and had the joint highest rating with the SASB-CMP (Schacht) formulation method, meaning that these two measures had the highest convergent validity with the other formulation methods[26]
Discriminant	No details
Factor structure	NA

continued

Responsiveness	
Discriminative (between individuals)	Differences in CCRTs for different diagnoses have not been established[17]
Evaluative (within individual across time)	The pervasiveness of the CCRT from the beginning to end of psychotherapy shows moderate consistency, with wish showing the greatest consistency[17]
	The same CCRT was identifiable in both early and later sessions. However, in the later sessions, the CCRT became more deeply experienced in the relationship with the therapist[16]
Acceptability	
Number of items	NA
Administration method	Judge-completed rating scale
Time taken to complete	Takes about 3 hours to score ten relationship episodes. Less time after practice
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1976, Plenum, New York
Web or scanning options	NA
Training details	Training details, worked scoring examples and case illustrations given in ref. 17. To be useful in reliability studies it is recommended that a candidate scorer should agree as high as a correlation of 0.75 with other judges
	Manuals available Latest edition of the *Guide to the CCRT method* (Luborsky & Crits-Christoph, 1998)
	Appendix A in ref. 17 demonstrates scoring and provides scoring symbols. Appendix B is a standard categories scoring table
	Ref. 16 includes a suggested format for the CCRT formulation. There is a chapter entitled 'Illustrations of the CCRT scoring guide' in ref. 20
Administration/process details	CCRT scoring method is in two phases. Phase A is for locating the RE and Phase B is for scoring the RE for the types of wishes, responses from other and responses of self
Support from measure developers	A set of sessions is being developed to serve for practising CCRT-based interpretations during played-back sessions (book, interrupted-session playback as practice in interpretation)
FAQ facility	No details
Precision	
Scale type	Each of the standard categories under the three components is scored on a five-point ordinal scale to rate intensity. Higher scores denote greater intensity
Normative data	16 cases (Luborsky, 1st ed., 6 October 1985) derived from psychotherapy sessions. All of the patients were in long-term psychotherapy: ten in psychoanalytic psychotherapy and six in psychoanalysis. The standard CCRT category list is an assemblage of those categories that best describe the core theme components expressed in the REs of the sample of 16 patients

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Résumé	
Strengths	The CCRT can be reliably applied; there is extensive support and there are training procedures for judges
	Method is clinically convenient and appropriate; method derived from analysis of therapy sessions
	CCRT useful guide to therapists in formulating interpretations
Weaknesses	CCRT method is quite time consuming and the training procedure is lengthy
	Method relies on access to tapes/transcripts of therapy sessions
Areas for further research	To simplify the method further
	To make the method more available to those who do not have access to narratives based on psychotherapy
	To develop further the method for distinguishing the more and the less conscious components of the CCRT
	To compare actual enactments of relationship events between patient and therapist in the session with the usual narratives told to the therapist about relationship events

Primary references

1. Albani C, Benninghofen D, Blaser G, Cierpka M, Dahlbender R, Geyer M, *et al*. On the connection between affective evaluation of recollected relationship experiences and the severity of the psychic impairment. *Psychother Res* 1999;**94**:452–67.
2. Barber JP, Luborsky L, Crits-Christoph P, Diguer L. A comparison of core conflictual relationship themes before psychotherapy and during early sessions. *J Consult Clin Psychol* 1995;**63**:145–8.
3. Bond JA, Hansell J, Shevrin H. Locating transference paradigms in psychotherapy transcripts: reliability of relationship episode location in the core conflictual relationship theme (CCRT) method. *Psychotherapy* 1987;**24**:736–49.
4. Bressi C, Amadei G, Astori S, Boato P, Colombo E, Coppola MT, *et al*. The therapeutic process in psychotherapy: a study of the Core Conflictual Relationship Theme. *New Trends Exp Clin Psychiatry* 1997;**13**:257–66.
5. Bressi C, Amadei G, Caparrelli S, Cattaneo C, Cova F, Crespi S, *et al*. A clinical and psychodynamic follow-up study of crisis intervention and brief psychotherapy in psychiatric emergency. *New Trends Exp Clin Psychiatr*, 2000;**16**,31–7.
6. Crits-Christoph P, Luborsky L, Dahl L, Popp C, Mellon J, Mark D. Clinicians can agree in assessing relationship patterns in psychotherapy: the Core Conflictual Relationship Theme method. *Arch Gen Psychiatry* 1988;**45**:1001–4.
7. Dazzi N, De Coro A, Ortu F, Andreassi S, Cundari M, Ostuni V, *et al*. The CCRT in an Italian sample of psychotherapies: a study on the 'wish' component. *New Trends Exp Clin Psychiatry* 1997;**13**:227–33.
8. Dazzi N, Petruccelli I. The 'Core Conflictual Relationship Theme (CCRT)' in an Italian sample of different psychotherapies. *New Trends Exp Clin Psychiatry* 1997;**13**:235–44.
9. Freni S, Azzone P. CCRT as a measure of psychotherapy process for two patients belonging to different diagnostic categories. *New Trends Exp Clin Psychiatry* 1997;**13**:245–56.
10. Hoglend P, Guldberg CA, Perry JC. Scientific approaches to making psychodynamic formulations. *Nord Psykiatr Tidsskr* 1992;**46**:41–8.
11. Johnson ME, Popp C, Schacht TE, Mellon J, Strupp HH. Converging evidence for identification of recurrent relationship themes: comparison of two methods. *Psychiatry* 1989;**52**:275–88.
12. Levine FJ, Luborsky L. The core conflictual relationship theme method: a demonstration of reliable clinical inferences by the method of mismatched cases. In Tuttman S, Kaye C, Zimmerman M, editors, *Object and self: a developmental approach*. New York: International Universities Press; 1981. pp. 501–26.
13. Luborsky L, Diguer L, Kachele H, Dahlbender R, Waldinger R, Freni S, *et al*. *A guide to the CCRT's methods, discoveries, and future*. Ulm Department of Psychotherapy and Psychosomatic Medicine; 1999. http://sip.medizin.uni-ulm.de/
14. Luborsky L, Kaechele H, Dahlbender R, Diguer L. *The CCRT Newsletter*. Philadelphia, PA: University of Pennsylvania; 1999.
15. Luborsky L. A pattern-setting therapeutic alliance study revisited. *Psychother Res* 2000,**10**:17–29.
16. Luborsky L. Measuring a pervasive psychic structure in psychotherapy: the Core Conflictual Relationship Theme. In Freedman N, Grand S, editors. *Communicative structures and psychic structures*. New York: Plenum Press; 1977. pp. 367–95.
17. Luborsky L. The Core Conflictual Relationship Theme: a basic case formulation method. In Eells TD, editor. *Handbook of psychotherapy case formulation*. New York: Guilford; 1997. pp. 58–83.
18. Luborsky L. New personality measures from old clinical concepts: a research agenda. In Routh DK, DeRubeis RJ, editors. *The science of clinical psychology: accomplishments and future directions*. Washington, DC: American Psychological Association; 1998. pp. 149–61.

continued

19. Luborsky L, Crits-Christoph P. A relationship pattern measure: the Core Conflictual Relationship Theme. *Psychiatry* 1989;**52**:250–59.
20. Luborsky L, Crits-Christoph P. *Understanding transference: the Core Conflictual Relationship Theme method*. New York: Basic Books; 1990.
21. Luborsky L, Crits-Christoph P, Friedman SH, Mark D, Schaffler P. Freud's transference template compared with the core conflictual relationship theme (CCRT): illustrations by the two specimen cases. In Horowitz MJ, editor. *Person schemas and maladaptive interpersonal patterns. The John D. and Catherine T. MacArthur Foundation series on mental health and development*. Chicago; IL: University of Chicago Press; 1991. pp. 167–95.
22. Luborsky L, Crits-Christoph P, Mellon J. Advent of objective measures of the transference concept. *J Consult Clin Psychol* 1986;**54**:39–47.
23. Luborsky L, Diguer L. A novel CCRT reliability study: reply to Zander *et al*. *Psychother Res* 1995;**5**:237–41.
24. Luborsky L, Crits-Christoph P, Friedman SH, Mark D, Schaffler P. A verification of Freud's grandest clinical hypothesis: the transference. *Clin Psychol Rev* 1985;**5**:231–46.
25. Luborsky L, Luborsky E. The era of measures of transference: the CCRT and other measures. In Shapiro T, Emde RN, editors. *Research in psychoanalysis: process, development, outcome*. New York: International Universities Press; 1995. pp. 329–51.
26. Luborsky L, Popp C, Barber JP. Common and special factors in different transference-related measures. *Psychother Res* 1994;**4**:277–86.
27. Luborsky L, Popp C, Luborsky E, Mark D. The Core Conflictual Relationship Theme. *Psychother Res* 1994;**4**:172–83.
28. Luborsky L, Crits-Christoph P. *Understanding transference: the Core Conflictual Relationship Theme method*. 2nd ed. Washington, DC: American Psychological Association; 1998.
29. Perry JC, Luborsky L, Silberschatz G, Popp C. An examination of three methods of psychodynamic formulation based on the same videotaped interview. *Psychiatry* 1989;**52**:302–23.
30. Popp C, Taketomo Y. The application of the core conflictual relationship theme method to Japanese psychoanalytic psychotherapy. *J Am Acad Psychoanal* 1993;**21**:229–52.
31. Wilczek A, Weinryb RM, Barber JP, Gustavsson JP, Asberg M. The Core Conflictual Relationship Theme (CCRT) and psychopathology in patients selected for dynamic psychotherapy. *Psychother Res* 2000;**10**:100–13.

Secondary references

32. Barber JP, Foltz C, DeRubeis RJ, Landis JR. Consistency of interpersonal themes in narratives about relationships. *Psychother Res* 2002;**12**:139–58.
33. Bennett D, Parry G. The accuracy of reformulation in cognitive analytic therapy: a validation study. *Psychother Res* 1998;**8**:84–103.
34. Book HE. *How to practice brief psychodynamic psychotherapy: the Core Conflictual Relationship Theme method*. Washington, DC: American Psychological Association; 1998.
35. Chance SSE, Bakeman R, Kaslow NJ, Farber E, Burge-Callaway K. Core conflictual relationship themes in patients diagnosed with borderline personality disorder who attempted, or who did not attempt, suicide. *Psychother Res* 2000;**10**:337–55.
36. Cierpka M, Strack M, Benninghoven D, Staats H, Dahlbender R, Pokorny D, *et al.* Stereotypical relationship patterns and psychopathology. *Psychother Psychosom* 1998;**67**:241–8.
37. Crits-Christoph P, Cooper A, Luborsky L. The accuracy of therapists' interpretations and the outcome of dynamic psychotherapy. *J Consult Clin Psychol* 1988;**56**:490–95.
38. Crits-Christoph P, Cooper A, Luborsky L. The accuracy of therapists' interpretations and the outcome of dynamic psychotherapy. In Hill CE, editor. *Helping skills: the empirical foundation*. Washington, DC: American Psychological Association; 2001. pp. 297–307.
39. Fried D, Crits-Christoph P, Luborsky L. The first empirical demonstration of transference in psychotherapy. *J Nerv Ment Dis* 1992;**180**:326–31.
40. Horowitz MJ. Relationship schema formulation: role–relationship models and intrapsychic conflict. *Psychiatry* 1989;**52**:260–74.
41. Luborsky L, editor. *The symptom-context method: symptoms as opportunities in psychotherapy*. Washington, DC: American Psychological Association; 1996.
42. Luborsky L. An introduction to central relationship pattern measures: the Central Relationship Questionnaire. *J Psychother Pract Res* 2000;**9**:200.
43. Luborsky L, Barber JP, Diguer L. The meanings of narratives told during psychotherapy: the fruits of a new observational unit. *Psychother Res* 1992;**2**:277–90.
44. Luborsky L. The benefits to the clinician of psychotherapy research: a clinician-researcher's view. In Talley PF, Strupp HH, Butler SF, editors. *Psychotherapy research and practice: bridging the gap*. New York: Basic Books; 1994. pp. 167–80.
45. Luborsky L, Crits-Christoph P, Mintz J, Auerbach A. *Who will benefit from psychotherapy? Predicting therapeutic outcomes*. New York: Basic Books; 1988.

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

46. Mitchell J. Coherence of the relationship theme: an extension of Luborsky's core conflictual relationship theme method. *Psychoanal Psychol* 1995;**12**:495–512.
47. Okey JL, McWhirter JJ, Delaney MK. The central relationship patterns of male veterans with posttraumatic stress disorder: a descriptive study. *Psychotherapy* 2000;**37**:171–9.
48. Popp CA, Diguer L, Luborsky L, Faude J, Johnson S, Morris M, *et al*. Repetitive relationship themes in waking narratives and dreams. *J Consult Clin Psychol* 1996;**64**:1073–8.
49. Staats H, May M, Herrmann C, Kersting A, Koenig K. Different patterns of change in narratives of men and women during analytical group psychotherapy. *Int J Group Psychother* 1998;**48**:363–80.
50. Turner RM. Launching cognitive-behavioral therapy for adolescent depression and drug abuse. In Budman SH, Hoyt MF, Friedman S, editors. *The first session in brief therapy*. New York; Guilford; 1992. pp. 135–55.

C15
Counseling Evaluation Inventory (CEI)

General details	
Authors	Linden JD, Stone SC, Shertzer B
Language	English
Country of publication/development	USA
Publication date	1965
Publisher	NA
Purpose and overview	
The CEI is a measure of counsellor effectiveness, and measures effectiveness on three dimensions of comfort, climate and satisfaction	
Theoretical orientation	Counselling
Population details	Adults with psychoneuroses, college students, secondary school or university students
Perspective	Client completed
Measure used by	Psychologists, counsellors, counsellor trainers/supervisors
Other versions	French version and French-Canadian version (Bachelor, 1987)
Notes	The CEI (21 items) was developed from a longer format CEI which had 68 items. The 21 items were selected if they loaded 0.40 or greater on one of the three factors (see Factor structure, below) and less than 0.40 on all other factors
Areas of therapist–patient interaction addressed: Map	
Therapy context: influence; power/coercion; responsibilities	
Roles: confidant	
Therapist engagement: empathy/sensitivity; respect; support/tolerance; listening	
Framework: collaborative/participative/Involving; congruent	
Therapeutic techniques: responsiveness/receptiveness/attunement	
Threats to the relationship: defensive; critical; hostility/anger	
Outcomes: general satisfaction; working alliance	
Information derived from items	
Dimensions	
Counseling climate	Eight items, e.g. the counsellor acted cold and distant
Counselor comfort	Five items, e.g. the counsellor gave the impression of feeling at ease
Client satisfaction	Seven items, e.g. the counsellor's discussion of test results was helpful to me
Reliability	
The CEI demonstrates adequate internal consistency and adequate test–retest reliability for all but one of its scales (which demonstrates partial reliability)	
Split-half	No details
Internal consistency	Average reliability coefficient (0.72)[3]
Inter-rater	NA
Test–retest	Test–retest coefficients: total scale: (0.83); counselling climate (0.78); counsellor comfort (0.63); client satisfaction (0.74)[3]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Validity	
The CEI demonstrated partial to adequate concurrent validity with the Counselor Rating Form (CRF) across scales	
Partial convergent validity with an outcome measure of perceived improvement was displayed	
There is mixed evidence regarding the three-factor structure of the CEI	
Face	See Content
Content	The initial 68 items were judged for social favourability: scores from 446 counsellors and 289 students showed no significant difference ($p > 0.05$) on these social favourability ratings, and were thus all retained[3]
Criterion (a) concurrent	Dimensions of the CEI were correlated with the expertness, attractiveness and trustworthiness dimensions of the CRF, as follows:[1] Expertness (CRF) and satisfaction: ($r = 0.60$) Expertness (CRF) and climate: ($r = 0.55$) Trustworthiness (CRF) and climate: ($r = 0.49$) Trustworthiness (CRF) and satisfaction: ($r = 0.53$) All $p < 0.001$ Expertness (CRF) and comfort: ($r = 0.45$) Attractiveness (CRF) and climate: ($r = 0.45$) Attractiveness (CRF) and satisfaction: ($r = 0.38$) All $p < 0.01$
Criterion (b) predictive	No details
Construct	The hypothesis that the CEI would be related to the CRF was supported[1] (see concurrent validity)
Convergent	The dimensions of the CEI correlated significantly with the outcome measure of perceived improvement ($p < 0.05$), with correlations 0.31–0.54[1]
Discriminant	No details
Factor structure	Using a sample of high-school students, a rotated factor matrix was applied and three factors (specified in dimensions) were found[3] The Linden *et al*. (1965) study[3] was replicated with college students instead of high-school students. The three-factor structure found by Linden *et al*. (1965) was not totally replicated as it appeared that the counsellor comfort dimension was a part of counselling climate. According to this study, significant intercorrelations among the subscales ($R = 0.36$ to 0.50) suggest that analysis of the subscales as independent constructs is unwarranted[2,4]
Responsiveness	
Discriminative (between individuals)	The CEI has been shown to discriminate between counselling trainees' final grades. On all three dimension scores all counsellors graded A in practicum were rated significantly higher ($p < 0.05$) than those rated grade C[3,4]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	21
Administration method	Self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	French
Access by ethnic minorities	The CEI has been used with black junior college students[6]

continued

Feasibility	
Copyright	1965, *Personnel and Guidance Journal*
Web or scanning options	No details
Training details	NA
Administration/process details	Completed by clients postcounselling. Can be mailed out to clients
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Five-point Likert-type scale. Higher scores indicate superior counsellor ratings
Normative data	No details
Résumé	
Strengths	Good internal consistency and test–retest reliability. Has been used with ethnically diverse samples
Weaknesses	Mixed evidence regarding the factor structure
Areas for further research	Further psychometric research to establish the factor structure of the CEI

Primary references

1. Bachelor A. The Counseling Evaluation Inventory and the Counselor Rating Form: their relationship to perceived improvement and to each other. *Psychol Rep* 1987;**61**:567–75.
2. Haase RF, Miller CD. Comparison of factor analytic studies of the counseling evaluation inventory. *J Counsel Psychol* 1968;**15**:363–7.
3. Linden JD, Stone SC, Shertzer B. Development and evaluation of an inventory for rating counseling. *Personnel Guid J* 1965;**44**:267–76.
4. Ponterotto JG, Furlong MJ. Evaluating counselor effectiveness: a critical review of rating scale instruments. *J Counsel Psychol* 1985;**32**:597–616.

Secondary references

5. Begley K, Ross MW, Austin P, Casey K, Collins P, Hennings G, *et al*. Development and evaluation of an inventory for rating client satisfaction with outcome in HIV counseling: the Albion Center Scale. *Patient Educ Counsel* 1994;**24**:341–5.
6. Brown RD, Frey DH, Crapo SE. Attitudes of black junior college students toward counseling services. *J Coll Stud Personnel* 1972;**13**:420–4.
7. Cormier LS, Hackney H, Segrist A. Three counselor training models: a comparative study. *Counsel Educ Supervis* 1974;**14**:95–104.
8. Goldfarb N. Effects of supervisory style on counselor effectiveness and facilitative responding. *J Counsel Psychol* 1978;**25**:454–60.
9. Graff RW. The relationship of counselor self-disclosure to counselor effectiveness. *J Exp Educ* 1970;**38**:19–22.
10. Heppner PP, Heesacker M. Perceived counselor characteristics, client expectations, and client satisfaction with counseling. *J Counsel Psychol* 1983;**30**:31–9.
11. Ivey AE, Miller CD, Gabbert KH. Counselor assignment and client attitude: a systematic replication. *J Counsel Psychol* 1968;**15**:194–5.
12. Markey MJ, Fredrickson RH, Johnson RW, Julius MA. Influence of playback techniques on counselor performance. *Counsel Educ Supervis* 1970;**9**:178–82.
13. Stockwell SR, Dye A. Effects of counselor touch on counseling outcome. *J Counsel Psychol* 1980;**27**:443–6.
14. Zarski JJ, Sweeney TJ, Barcikowski RS. Counseling effectiveness as a function of counselor social interest. *J Counsel Psychol* 1977;**24**:1–5.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C16
Counselor Effectiveness Rating Scale (CERS)

General details	
Author	Atkinson DR
Language	English
Country of publication/development	USA
Publication date	1982
Publisher	NA
Purpose and overview	
To measure a client's or an observer's perception of counsellor effectiveness through assessment of a counsellor's social influence behaviours; this measure shares its purpose with the Counselor Rating Form (CRF; Barak & LaCrosse, 1975)	
Theoretical orientation	No details
Population details	See notes
Perspective	Independent rater/therapist rated by client
Measure used by	Researchers, clinicians
Other versions	Initially developed in 1975 by Atkinson and Carskaddon,[4] and modified by Furlong, Atkinson and Caas in 1979. Studies that used these earlier versions report no validity or reliability evidence
Notes	*Practitioners:* Rogers in *Three approaches to psychotherapy*[1] Counsellors with varying degrees of strength along the measure dimensions[3] Counsellor high in prestige or using jargon[4] Mexican-American vs Anglo-American counsellor[7] Indian vs non-Indian counsellor and trustworthy vs untrustworthy[8] *Clients:* Depressed freshman[3] Clients from drug abuse programmes[4] *Independent raters:* Introductory psychology students[6] Students[1,3,4] Mexican-American high-school students[5,7] American Indian high-school students[8]
Areas of therapist–patient interaction addressed: Map	
Roles: expert/authority/leader	
Therapist engagement: genuineness	
Patient engagement: attraction	
Achieving a working relationship: working alliance affective bond (trust, liking, caring)	
General satisfaction: satisfaction	
Inferred from a brief description	
Dimensions	
Expertness	Three items
Trustworthiness	Three items
Attractiveness	Three items
Counsellor utility	One item: 'Someone I would see for counselling'

continued

Reliability	
Internal consistency as estimated by Cronbach's alpha was adequate in all reported studies. No other areas of reliability were addressed	
Split-half	No details
Internal consistency	Expertness 0.88, trustworthiness 0.75, attractiveness 0.78[1]
	Atkinson and Wampold (1982) alpha coefficients were: expertness 0.88; trustworthiness 0.75; attractiveness 0.78; total 0.90[2]
	Expertness 0.88, trustworthiness 0.85, attractiveness 0.87[3]
Inter-rater	No details
Test–retest	No details
Validity	
Concurrent validity with CERS was adequate both when looking at total and subscale scores. Concurrent validity ranged from partial to adequate for CRF and CERS correlations for different dimension pairings. Adequate predictive validity and convergent validity with the CERS has been demonstrated	
Evidence of discriminant validity was poor. Factor structure findings are mixed	
Face	No details
Content	No details
Criterion (a) concurrent	Atkinson and Wampold (1982) correlated CERS scores with Counsellor Rating Form (CRF) scores and found a validity coefficient of 0.80 for the CERS[1,2]
	Correlation between total CERS score and total CRF = 0.80. For correlations between dimensions of both scales, expertness (0.79), trustworthiness (0.73) and attractiveness (0.73); these are large enough to indicate convergent validity ($p < 0.10$)
	CRF:CERS correlations for different dimension pairings varied 0.42–0.79[1]
	The CERS correlates with the CRF and CRF-S for each of the shared dimensions at $p < 0.01$: for expert (0.83–0.86), for attractive (0.80–0.87), for trustworthy (0.83–0.86)[3]
Criterion (b) predictive	Atkinson and Wampold (1982)[1] found the CERS predictive of willingness to self-refer for counselling ($r = 0.67$)[1,2]
Construct	No details
Convergent	Atkinson and Wampold (1982)[1] found convergent coefficients of 0.73 to 0.79 for the CERS and CRF[2]
Discriminant	This was interpreted as not existing, owing to some of the correlations between different dimensions based on the same scale being larger than some of the same dimension correlations across scales[1]
	Intercorrelations between the three dimensions were higher than would be desired (0.54–0.77)[3]
Factor structure	A maximum likelihood factor analysis with varimax orthogonal rotation was conducted. Poor replication of the intended factor structure was found. While all of the expertness items loaded optimally on to the first factor, there was no clear pattern of loadings of the trustworthiness and attractiveness items on the second and third factors[1]
	Maximum likelihood factor analysis with varimax orthogonal rotation was used by Atkinson and Wampold (1982).[1] The percentages of total variance accounted for by expertness, attractiveness and trustworthiness were 29.0, 16.5 and 21.1 respectively, a total of 66.6[2]
	Principal component analysis, utilising factors with eigenvalues >1, with orthogonal rotation, produced a single, general evaluative factor accounting for 62% of variance on OBLIMIN rotation for principal components analysis for the three measures combined; while eight factors emerged, the pattern of optimal factor loadings was more suggestive of a three-factor solution, which would be consistent with the elements of social influence theory[3]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	Ten
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1982, *Counselor Education and Supervision*
Web or scanning options	No details
Training details	No details
Administration/process details	Administered following viewing of videotaped interactions
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert
Normative data	Overall mean: expert 9.7 (SD 4.2), attractive 8.8 (SD 4.2) and trustworthy 11.1 (SD 4.7)[3]
Résumé	
Strengths	Adequate internal consistency, concurrent, predictive and convergent validity have been demonstrated
Weaknesses	Poor discriminant validity
	Researchers and clinicians using this scale need to be aware that client responses may be heavily influenced by a 'good person' or 'cooperative subject' response bias (Bergin, 1971)[2]
	Findings suggest that the CRF is not only subject to a strong ceiling effect but also insensitive to the effects of varying levels of counsellor expertise[2]
Areas for further research	Factor structure in an attempt to resolve mixed findings

Primary references

1. Atkinson DR, Wampold BE. A comparison of the Counselor Rating Form and the Counselor Effectiveness Rating Scale. *Counsel Educ Supervis* 1982;**22**:25–36.
2. Ponterotto JG, Furlong MJ. Evaluating counselor effectiveness: a critical review of rating scale instruments. *J Counsel Psychol* 1985;**32**:597–616.
3. Wilson FR, Yager GG. Concurrent and construct validity of three counselor social influence instruments. *Measure Eval Counsel Dev* 1990;**23**:52–66.

continued

Secondary references

4. Atkinson DR, Carskaddon G. A prestigious introduction, psychological jargon, and perceived counselor credibility. *J Counsel Psychol* 1975;**22**:180–6.
5. Atkinson DR, Ponce FQ, Martinez FM. Effects of ethnic, sex, and attitude similarity on counselor credibility. *J Counsel Psychol* 1984;**31**:588–90.
6. Barak A, LaCrosse MB. Multidimensional perception of counselor behavior. *J Counsel Psychol* 1975;**22**:471–6.
7. Hess RS, Street EM. The effect of acculturation on the relationship of counselor ethnicity and client ratings. *J Counsel Psychol* 1991;**38**:71–5.
8. LaFromboise TD, Dixon DN. American Indian perception of trustworthiness in a counseling interview. *J Counsel Psychol* 1981;**28**:165–9.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C17
Counsellor Effectiveness Scale (CES)

General details	
Author	Ivey AE
Language	English
Country of publication/development	UK
Publication date	1971
Publisher	NA
Purpose and overview	
For use in microcounselling training and research to measure client-perceived counsellor effectiveness	
Theoretical orientation	Pan-theoretical/counseling psychology
Population details	See below
Perspective	Independent rater
Measure used by	Research psychologists/counselling psychologists/counsellors
Other versions	Two parallel forms constructed of 25 items each
Notes	Used in vocational counselling context[4] *Raters:* Undergraduates[2] 22–35-year-old counselling trainees[5]
Areas of therapist–patient interaction addressed: Map	
Therapy context: responsibilities	
Therapist engagement: empathy/sensitivity; listening; genuineness; praise/affirmation; warmth	
Roles: expert/authority/leader; confidant	
Therapeutic techniques: responsiveness/receptiveness/attunement	
Threats to the relationship: fear; resistance; critical; hostility/anger	
Outcomes: safety/secure base; affective bond; satisfaction	
Framework: focused; structuring	
Information derived from items in ref. 2	
Dimensions	
Positiveness	
Calmness	
Animation	
Reliability	
Very little reliability information. Inadequate inter-rater reliability	
Split-half	No details
Internal consistency	No details
Inter-rater	In an Ivey and Authier (1978) study, undergraduate judges made 50 observations from a videotape model. Inter-rater reliability was significant, with Kendall $w = 0.37$ ($p < 0.001$)[3]
Test–retest	No details

continued

Validity	
Very little validity information has been derived from use of the CES.	
Face	No details
Content	The 25 items were selected on the basis of 30 graduate students rating two models of counselling, one desirable and the other undesirable or ineffective, using a pool of 93 items[1]
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	For the most part, the CES has been used as a unitary measure of perceived counsellor effectiveness. However, Nay (in Negata *et al*., 1983) used a principal components analysis with CES data and three consistent components with factor loadings of ⩾0.55 emerged (positiveness, calmness and animation)[3]
Responsiveness	
Discriminative (between individuals)	Ivey and Authier (1978) tested the CES's discriminate validity by having undergraduates rate one effective and one ineffective counsellor on the two CES forms. Two *t*-tests yielded highly significant ($p < 0.001$) differences between the two counsellors' ratings using both forms[3]
Evaluative (within individual across time)	Ivey and Authier (1978) found the CES to be highly reactive to changes in the client's environment, and they caution against its use in situations other than immediate pre- or post-training microcounselling sessions[3]
Acceptability	
Number of items	25
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1971
Web or scanning options	No details
Training details	No details
Administration/process details	No details
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert. Seven-point semantic differential (1–7).
Normative data	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Notes	
The CES is also known as the Rating Scale of Counsellor Effectiveness (RSCE) A parallel form (representing positive and negative counselling models) of reliability (coefficiency of equivalence) was computed for the two forms by Ivey and Authier (1978). Psychology students rated videotaped counselling and the parallel form reliability was 0.98[3]	
Résumé	
Strengths	The CES has been primarily used to measure client attitudes towards their counsellor. When used in evaluating the counsellor before and after microtraining sessions, it has proven to be a sensitive and useful instrument (Ivey, 1979). Demonstrated responsiveness
Weaknesses	However, the CES has also been found to be highly reactive to changes in the client's atmosphere and thus its use outside immediate pre- and post-training is not recommended (Ivey, 1979) Limited psychometric validation
Areas for further research	Further testing of psychometric properties

Primary references

1. Ivey AE. *Microcounseling: innovations in interview training*. Springfield, IL: CC Thomas; 1971.
2. Ivey AE, Authier J. *Microcounseling*. 2nd ed. Springfield, IL: CC Thomas; 1978.
3. Ponterotto JG, Furlong MJ. Evaluating counselor effectiveness: a critical review of rating scale instruments. *J Counsel Psychol* 1985;**32**:597–616.

Secondary references

4. Guttman MA, Haase RF. Effect of experimentally induced sets of high and low 'expertness' during brief vocational counseling. *Counsel Educ Supervis* 1972;**11**:171–8.
5. Malikiosi-Loizos M, Gold JA, Mehnert WO, Work GG. Differential supervision and cognitive structure effects on empathy and counseling effectiveness. *Int J Advance Counsell* 1981;**4**:119–29.

C18
Counselor Rating Form (CRF)

General details	
Authors	Barak A, LaCrosse MB
Language	English
Country of publication/development	USA
Publication date	1975
Publisher	NA
Purpose and overview	
The Counselor Rating Form (CRF) is a measure of client- or observer-perceived counsellor behaviour. It was designed as a means of investigating Strong's (1968; see ref. 3) prediction of the existence of three dimensions of perceived counsellor behaviour: expertness, attractiveness and trustworthiness	
Theoretical orientation	Social psychology, specifically social influence theory
Population details	Clinical adults, clinical adolescents and non-clinical adults
Perspective	The CRF is completed by the client or an independent rater
Measure used by	Researchers
Other versions	The CRF has French and Short versions, which are included in this review
Notes	The sample used in the development of the CRF consisted of 202 psychology students who rated Rogers, Ellis and Perls in the film *Three approaches to psychotherapy*[3]
Areas of therapist–patient interaction addressed: Map	
Therapy context: confidential; values; responsibilities	
Roles: confidant; expert/authority/leader; good object	
Therapist engagement: warmth; genuineness; respect; openness	
Patient engagement: attraction	
The therapist–client interaction information is derived from the scale items	
Dimensions	
Expertness	12 items relating to perceived counsellor expertness
Attractiveness	12 items relating to perceived counsellor attractiveness
Trustworthiness	12 items relating to perceived counsellor trustworthiness
Reliability	
The CRF is subject to a ceiling effect[9]	
The internal consistencies and split-half reliabilities of each of the three dimensions have been assessed in four studies and shown to be adequate	
75% agreement among four judges as to the appropriate category (expertness, attractiveness and trustworthiness) of an initial 83 items was the criterion for the CRF item selection. Therefore, all items have at least 75% inter-rater agreement[3]	
Split-half	Spearman–Brown correlation coefficients for expertness, attractiveness and trustworthiness; 0.87, 0.85 and 0.90, respectively (*p*-values not reported)[3]
Internal consistency	Alpha coefficients are reported for expertness, attractiveness and trustworthiness respectively as follows: 0.82, 0.91 and 0.98[1] 0.94, 0.90 and 0.91[10] 0.77 to 0.97, mean 0.86; 0.83 to 0.92, mean 0.88; and 0.86 to –0.91, mean 0.88[5]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Inter-rater	Of the CRF's 36 items, 22 and 14 attained 100% and 75% agreement, respectively, from four judges as to their appropriate category (expertness, attractiveness and trustworthiness)[3]
Test–retest	No details
Validity	

Face and content validity
The CRF has face and content validity in that the 36 items were selected by expert judges, and met the criterion of having agreement from 75% of the judges as to its appropriate category (expertness, attractiveness or trustworthiness)[3]

Predictive validity
CRF scores (total and for each dimension) have been shown to predict outcome as assessed by Goal Attainment Scaling (GAS; see ref. 7). This is the case for CRF scores taken before counselling and after counselling.[7] The CRF has predicted client satisfaction as measured by the Client Evaluation Inventory (CEI; Linden *et al*., 1965, see ref. 6) as both a one-factor and a three-factor model.[6] The CRF has demonstrated partial and adequate convergent validity with client-perceived outcome[2]

Construct validity
In an examination of the construct validity of the CRF (in that it measures three distinct traits), two hypotheses were tested. The first was that correlations between equivalent dimensions on the CRF and Counselor Evaluation Rating Scale (CERS; see ref. 1) scales would be higher than correlations between different dimensions on the two scales (e.g. that CRF expertness and CERS expertness would correlate more highly than CRF expertness and CERS attractiveness). The second hypothesis was that equivalent dimensions *between* the two scales would correlate more highly than dimensions *within* each scale (e.g. CRF expertness and CERS expertness would correlate more highly than CRF expertness and CRF attractiveness). The first hypothesis was supported, while the second was not[1]

Convergent validity
CRF total scores and each of the dimensions have demonstrated adequate convergent validity with the total scores and equivalent dimensions[1]

Factor structure
The factor structure of the CRF has been assessed with factor analyses and by intercorrelating the three dimensions. The results have been mixed, with some suggesting a three-factor model, while others suggest that the dimensions are not so distinct

Also, with regard to factor structure, the CRF has been shown to predict client satisfaction as measured by the CEI equally well as a one-factor or a three-factor model (see also Predictive validity)[6]

Face	See Content validity
Content	Four expert judges selected the 36 items from a list of 83. Each of the 83 items was classified into one of the three dimension categories and the 36 items were selected on the basis of at least 75% agreement on which category the item belonged to[3]
Criterion (a) concurrent	No details
Criterion (b) predictive	Correlation coefficients between precounselling CRF and GAS (outcome) scores were as follows (n = 36): Total CRF score 0.53 ($p < 0.001$), expertness 0.56 ($p < 0.001$), attractiveness 0.45 ($p < 0.01$) and trustworthiness 0.37 ($p < 0.01$)[7] The same research reports the same pattern, but larger correlations for postcounselling CRF and GAS scores[7] Stepwise multiple regression analysis showed that the three dimensions together accounted for 35.2% of the variance on outcome ($R = 59.4$, $F_{3,32} = 5.81$, $p < 0.01$) and expertness alone accounted for 31.1% ($R = 0.558$, $F_{1,34} = 15.38$, $p < 0.001$) ($n = 36$)[7] Correlations between CRF scores and client-perceived outcome ranged from 0.33 to 0.53 ($p < 0.05$)[2] Correlations between CRF dimensions and willingness to self-refer ranged from 0.23 to 0.67, with a median of 0.47[3]

continued

	A one-factor CRF predicted 56 out of 86 clients' satisfaction (CEI) scores, with an error rate of 22%, where chance predictions would produce an error rate of approximately 50%[6]
	The results of a multiple regression analysis indicated that the one-factor CRF model, with openness, accounted for 74% of the variance in CEI (outcome) scores. CRF was a significant predictor ($F_{2,71} = 176.08$, $p < 0.0001$), whereas openness was not[6]
	Multiple regression analysis of the contribution of openness and the three CRF dimensions to CEI (outcome) scores indicated that the model accounted for 75% of the total variance, and expertness was the only significant predictor ($F_{4,71} = 20.10$, $p < 0.0001$)[6]
	As hypothesised, equivalent dimensions across CRF and CERS scales had higher correlations than did different dimensions across the two scales (e.g. CRF expertness and CERS expertness correlated more highly than CRF expertness and CERS attractiveness), ($z = 1.90$, $p < 0.05$)[1]
Construct	The second hypothesis was not supported as correlations among dimensions within each scale ($r = 0.69$) were not significantly lower than correlations among equivalent dimensions between the CRF and CERS ($r = 0.75$), ($z = 0.64$)[1]
Convergent validity	Correlation coefficients between CRF expertness, attractiveness and trustworthiness with CERS equivalent scales were 0.80, 0.79, 0.73 and 0.73, respectively (p-values not reported)[1]
	CRF and CERS equivalent dimensions were correlated using the Hubert and Baker (1978; see ref. 1) procedure produced a z-value of 1.60 ($p < 0.10$)[1]
Discriminant validity	No details
Factor structure	Principal factor analysis with varimax rotation was conducted on ratings of Rogers, Ellis and Perls with the following results:
	For the ratings of Rogers, all Expertness items loaded onto Factor 1, all Attractiveness items loaded onto Factor 2, four of which cross loaded and nine out of 12 Trustworthiness items loaded onto Factor 3, eight of which cross loaded[4]
	For the ratings of Perls, all of the Expertness items loaded onto Factor 1, while the patterns is unclear for the Attractiveness and Trustworthiness items[4]
	The analysis of the ratings for Ellis produced no clear patterns[4]
	The CRF was subject to principal components analyses with varimax rotation and the dimension items were distributed as follows:
	11 of the 12 expertness items loaded onto factor 2 with eigenvalues ranging from 0.50 to 0.84[10]
	Ten of the 12 attractiveness items loaded onto factor 1, with loadings ranging from 0.49 to 0.79[10]
	Nine of the 12 trustworthiness items loaded onto factor 3, with eigenvalues ranging from 0.50 to 0.71[10]
	When the CRF items were subjected to OBLIMIN rotation, they were distributed as follows:
	Ten of the 12 attractiveness items loaded onto factor 1, with loadings ranging from 0.52 to 0.88[10]
	Eight of the 12 trustworthiness items loaded onto factor 3, with eigenvalues ranging from 0.56 to 0.74[10]
	Eleven of the 12 expertness items loaded onto factor 2, with eigenvalues ranging from 0.54 to 0.87[10]
	These results of the OBLIMIN rotation suggest two main factors: attractiveness–trustworthiness and expertness[10]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	In a maximum likelihood factor analysis with varimax orthogonal rotation, all of the expertness items loaded onto the first factor (loadings from 0.39 to 0.85). The trustworthiness and attractiveness items produced no clear patterns on the second and third factors[1] Principal components factor analysis resulted in five factors with eigenvalues over 1, accounting for 71% of the common variance, with factor 1 alone accounting for 56%[6] The CRF has similar predictive validity as either a one-factor or three-factor model (see Predictive validity, also ref. 6) Intercorrelations among expertness, attractiveness and trustworthiness have ranged from: 0.54 to 0.77[10] 0.75 to 0.93 (median 0.69)[8] 0.82 to 0.98[1] 0.30 to 0.92 (median 0.77)[5] Using Dunn's (1961, see ref. 8) post hoc pairwise comparison procedure, some significant differences between dimensions were found in the individual ratings of Rogers, Perls and Ellis (e.g. Perls' ratings of attractiveness were higher than his ratings of trustworthiness ($p < 0.01$)[8]
Responsiveness	
Discriminative (between individuals)	Average ratings of counsellors with at least 3 years of experience were not consistently higher than ratings of people with no formal counselling experience[3] Using Dunn's (1961, see ref. 8) post hoc pairwise comparison procedure, some significant differences were found in ratings of Rogers, Perls and Ellis (e.g. Perls and Ellis were perceived as more expert than Rogers, $p < 0.01$)[8]
Evaluative (within individual across time)	A significant ($p < 0.01$) main effect has been found in ratings of Rogers, Perls and Ellis[3] t-Tests have shown a significant increase from pre- to post-counselling CRF ratings as follows: Expertness, $t = 2.74$ ($p < 0.001$) Attractiveness, $t = 2.32$ ($p < 0.05$) Trustworthiness, $t = 2.26$ ($p < 0.05$) Total, $t = 2.89$ ($p < 0.001$)[7]
Acceptability	
Number of items	36
Administration method	Rating scale
Time taken to complete	Average time of 12.5 minutes reported[22]
Flesch reading age	Reading level of 12th grade
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1975, *Journal of Counseling Psychology*
Web or scanning options	No details
Training details	No details on the training of raters. A confederate client has been trained in one study[3]
Administration/process details	Independent raters rate the counsellor after viewing video footage of the session. Client raters are instructed to complete the form immediately after the session

continued

Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Likert-type. Each item consists of a pair of bipolar adjectives (e.g. reliable–unreliable) rated on seven-point scales
Normative data	Normative data available in refs 4, 7 and 8
Notes	

The CRF items are listed in ref. 3 and modifications are reported in ref. 8

Other uses of the CRF and its theoretical concepts in research include:

An assessment of the effects of practitioners' professional affiliation (social work, psychiatry and psychology), gender and warmth on participants' perceptions of the expertise and social attractiveness of the three types of practitioner[11]

A study with male undergraduates of the extent to which a counsellor quality (reputed expertise) and a client quality (self-concept) were interactive with counsellor gender and with each other as determinants of subjects' perceptions of the counsellor[12]

A study with homosexual men investigating perceived counsellor credibility and attractiveness as a function of sexual preference similarity and attitude similarity (gay advocacy) between rater and counsellor[13]

A repeated measures analysis comparing three sources (clients, counsellors, and supervisors) of ratings of counsellor expertise, attractiveness and trustworthiness[14]

A study with 120 undergraduates, in a factorial design, examining the effects and the relative contribution of three counsellor behaviours (non-verbal behaviour, jargon and attire) on perceived expertness and attractiveness[15]

A study with undergraduates to test the hypotheses that (1) counsellors using particular verbal interventions (interpretation or restatement) and non-verbal behaviour (high or low responsiveness) would be perceived to be differentially expert, trustworthy and attractive, and (2) they would be differentially able to influence[16]

A comparison of the effects of computer-mediated online counselling (via the Internet) and traditional face-to-face counselling on anxiety and attitudes towards counselling using 24 undergraduates as clients and six male graduate students as counsellors[17]

A study with 60 female undergraduates that examined the effects of a counsellor's fee level and title on perceptions of counsellor behaviour[18]

An analogy study investigating the perceptions of 75 undergraduate women regarding paraprofessional and professional therapists[19]

A social influence model of supervision is outlined based on research on interpersonal influence in counselling. In the model, the three general counsellor characteristics of perceived expertness, trustworthiness and attractiveness are shown to contribute to the supervisor's social power base[20]

A study with 116 undergraduate counselling clients that indicated that clients with a high degree of motivation who perceived their counsellors as persons with a high degree of social power improved in self-concept over the course of counselling. Counsellor trustworthiness and client motivation were predictive of change in self-concept, while counsellor expertise and attractiveness were not[21]

A study with 120 undergraduates to determine whether there are differences between the CRF and CRF-S in assessing client perceptions of the counsellor, and time taken to complete[22]

A study with 84 undergraduates, which investigated client willingness to refer to a counsellor as a function of the counsellor's gender and matching client predicates (visual, auditory or kinaesthetic)[23]

An examination of the effect of six counsellor verbal responses on clients' verbal behaviour and on their perceptions of counsellors[24]

An examination of Bandler and Grinder's (1976) statement that trust in a relationship will be enhanced if the counsellor matches the client's primary representational system[25]

An investigation into the effects of counsellor gender and gender role and client gender and presenting problem on CRF scores[26]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

A study with 161 undergraduates to investigate the relationship between participants' perceptions of counsellor expertness, trustworthiness and attractiveness and the level of facilitative and action dimensions displayed by the counsellor[27]

A study that investigated the effect of 12 descriptions of the counsellor's training, experience and similarity on the perceptions of the counsellor by 96 hearing-impaired college students and on their willingness to see the counsellor[28]

An examination of the relationship of social influence variables, symptom change and premature termination of counselling in 51 adult outpatients in counselling[29]

A field study that examined the relationships among a client's gender role attitude, the client's gender, the counsellor's gender and the client's rating of his or her counsellor[30]

An examination of the interpersonal influence process within an actual counselling context over an average of eight sessions[31]

A study that examined (1) the relationship between perceived counsellor expertness, attractiveness and trustworthiness and client satisfaction; (2) the relationships between specific client expectations on perceived counsellor characteristics and client satisfaction; and (3) the effects of actual counsellor experience level on perceived counsellor characteristics and client satisfaction[32]

An investigation into counsellor touch in the initial counselling session. Participants were in either a touch or no-touch condition and completed, among other measures, the CRF[33]

An investigation of similarities and differences between 83 college students' existing conceptions of counsellor characteristics and behaviours and their subsequent perceptions of these same characteristics and behaviours following videotaped samples of two counselling interactions that demonstrated client-centred and rational-emotive therapy[34]

A study that investigated the types of therapeutic variables considered by observers to both differentiate between, and contribute to, clients' positive change in behavioural and insight-orientated therapies[35]

Examples of research using the CRF:

A study that examined 160 college students' reactions to a therapist's and/or client's use of profanity (PF), using the CRF and the Self-referral Questionnaire[36]

A study that examined the effects of specific verbal and non-verbal behaviours on initial evaluations of counsellors[37]

An investigation into the comparative perceptions of counsellor behaviour[38]

A counselling analogue study to evaluate the effects of counsellor trustworthiness and counsellor ethnicity on Native American student ratings of perceived counsellor trustworthiness[39]

A study comparing the effects of counsellor self-disclosure vs counsellor self-involving statements on ratings of counsellors' expertness, attractiveness and trustworthiness (CRF)[40,41]

A study to determine the effects of counsellor status (high, low), counsellor weight (normal, overweight) and client gender on initial perceptions of counsellor expertness, attractiveness and trustworthiness[42]

A study into counsellor breach of client confidentiality and observer-rated counsellor trustworthiness. Among other measures participants rated counsellors on trustworthiness on the CRF[43]

A study that tested the effects of counsellor interpretation style, summary statements and restatements on perceived counsellor social influence and willingness to see the counsellor[44]

An investigation of the effects of sexual orientation similarity of counsellor and client as well as counsellor experience level on perceptions of counsellors by gay men and lesbians[45]

A study of clinical psychologists', counselling psychologists', psychiatrists' and social workers' ratings of their own and each of the other groups along three variables (attractiveness, expertise and trustworthiness) using the CRF[46]

A study that investigated the ability of 75 graduate counsellor trainees to recognise gender bias in client–counsellor interactions and examined how two different sets of instructions given to each participant influenced awareness of gender bias and perceptions of a counsellor[47]

An investigation into the role of verbal and non-verbal cues in the formation of first impressions of black and white counsellors[48]

A study in which expert and referent power bases and influence attempts were crossed with levels of trustworthiness to explore the effects of perceived counsellor illegitimacy and power base influence attempt incongruence in a counselling analogue[49]

continued

A study that examined the effects of neurolinguistic mirroring vs non-mirroring of selected non-verbal behaviours on empathy, trustworthiness and positive interaction in a cross-cultural setting among 60 Choctaw male adolescents and two white female counsellors[50]

A study in which undergraduates viewed a simulated counselling tape where the approach to counselling was either consistent with or discrepant from the client's cultural norms and values. Participants then rated the counsellor using the CRF[51]

An examination of the influence of client–counsellor group membership similarity, counsellor reputational cues and counsellor attending behaviour on disabled clients' perceptions of counsellor's attractiveness and expertness[52]

A study that investigated the effect of gender-role-incongruent behaviour on evaluations of counsellor expertness, attractiveness and trustworthiness[53]

A study that examined how counsellor race influences client evaluation of counselling effectiveness[54]

A study that tested the prediction that participant observers would give higher ratings on the CRF to self-disclosing counsellors than to non-disclosing ones. Higher ratings were given to self-disclosing counsellors[55]

A study that investigated client characteristics and counsellor perceptions. The paper suggests that the CRF can be used to obtain counsellor as well as client perceptions[56]

An examination of the hypothesis that positive self-involving and self-disclosing counsellor responses would be rated more favourably by participants than negative self-involving and self-disclosing counsellor responses[57]

A field study on the social influence process in counselling[58]

Résumé

Strengths	The form can be completed in approximately 12 minutes. 75% agreement among four judges as to the appropriate category (expertness, attractiveness and trustworthiness) of an initial 83 items was the criterion for the CRF item selection. Each dimension has adequate split-half reliability.[3] The internal consistency of each of the three dimensions was consistently adequate across three studies.[1,5,10] It has been assessed in three studies and shown to be adequate CRF scores have predicted outcome in four studies.[2,3,6,7] The hypothesis that equivalent dimensions on the CRF and Counselor Evaluation Rating Scale (CERS; see ref. 1) would correlate more highly than different dimensions between the two scales was supported[1] Each of the dimensions, and total scores have demonstrated adequate convergent validity with the CERS[1] The CRF is responsive to the perceived differences between Rogers, Ellis and Perls, in the film *Three approaches to psychotherapy*.[3,8] The form is also sensitive to perceived differences in individual counsellors from pre- to post-therapy[7]
Weaknesses	Contrary to expectations, equivalent dimensions *between* CRF and CERS correlated at lower levels than dimensions *within* each scale.[1] CRF total scores did not adequately predict client-perceived outcome[2]
Areas for further research	Further assessment of psychometric properties, e.g. convergence with other measures of counsellor qualities

Primary references

1. Atkinson DR, Wampold BE. A comparison of the Counselor Rating Form and the Counselor Effectiveness Rating Scale. *Counsel Educ Supervis* 1982;**22**:25–36.
2. Bachelor A. The Counseling Evaluation Inventory and the Counselor Rating Form: their relationship to perceived improvement and to each other. *Psychol Rep* 1987;**61**:567–75.
3. Barak A, Del DM. Differential perceptions of counselor behavior: replication and extension. *J Counsel Psychol* 1977;**24**:288–92.
4. Barak A, LaCrosse MB. Multidimensional perception of counselor behavior. *J Counsel Psychol* 1975;**22**:471–6.
5. Epperson DL, Pecnik JA. Counselor Rating Form – Short Version: further validation and comparison to the long form. *J Counsel Psychol* 1985;**1**:3–146.
6. Heesacker M, Heppner PP. Using real-client perceptions to examine psychometric properties of the Counselor Rating Form. *J Counsel Psychol* 1983;**2**:80–187.

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

7. LaCrosse MB. Perceived counselor social influence and counseling outcomes: validity of the Counselor Rating Form. *J Counsel Psychol* 1980;**27**:320–7.
8. LaCrosse MB, Barak A. Differential perception of counselor behavior. *J Counsel Psychol* 1976;**23**:170–2.
9. Ponterotto JG, Furlong MJ. Evaluating counselor effectiveness: a critical review of rating scale instruments. *J Counsel Psychol* 1985;**32**:597–616.
10. Wilson FR, Yager GG. Concurrent and construct validity of three counselor social influence instruments. *Measure Eval Counsel Dev* 1990;**23**:52–66.

Secondary references

11. Alperin RM, Neidengard TH. Effects of practitioners' professional affiliation, sex, and warmth on changes in the attitudes of clients. *Soc Work Res Abstr* 1984;**20**:20–26.
12. Angle SS, Goodyear RK. Perceptions of counselor qualities: impact of subjects' self-concepts, counselor gender, and counselor introductions. *J Counsel Psychol* 1984;**31**:576–9.
13. Atkinson DR, Brady S, Casas JM. Sexual preference similarity, attitude similarity, and perceived counseling credibility and attractiveness. *J Counsel Psychol* 1981;**28**:504–9.
14. Barak A, Lacrosse MB. Comparative perceptions of practicum counselor behavior. *Counsel Educ Supervis* 1977;**16**:202–8.
15. Barak A, Patkin J, Dell DM. Effects of certain counselor behaviors on perceived expertness and attractiveness. *J Counsel Psychol* 1982;**29**:261–7.
16. Claiborn CD. Counselor verbal intervention, nonverbal behavior, and social power. *J Counsel Psychol* 1979;**26**:378–83.
17. Cohen GE, Kerr BA. Computer-mediated counseling: an empirical study of a new mental health treatment. *Comput Hum Serv* 1998;**15**:13–26.
18. Conoley JC, Bonner M. The effects of counselor fee and title on perceptions of counselor behavior. *J Counsel Dev* 1991;**69**:356–8.
19. Corcoran KJ. Unraveling subjects' perceptions of paraprofessionals and professionals: a pilot study. *Percept Motor Skills* 1985;**60**:111–14.
20. Dorn FJ. Utilizing the social influence model in clinical supervision. *Clin Supervis* 1985;**3**:77–84.
21. Dorn FJ, Day BJ. Assessing change in self-concept: a social–psychological approach. *Am Ment Health Counsel Assoc J* 1985;**7**:180–86.
22. Dorn FJ, Jereb R. Enhancing the usability of the Counselor Rating Form for researchers and practitioners. *Measure Eval Counsel Dev* 1985;**18**:12–16.
23. Dowd ET, Pety J. Effect of counselor predicate matching on perceived social influence and client satisfaction. *J Counsel Psychol* 1982;**29**:206–9.
24. Ehrlich RP, D'Augelli AR, Danish SJ. Comparative effectiveness of six counselor verbal responses. *J Counsel Psychol* 1979;**26**:390–8.
25. Falzett WC. Matched versus unmatched primary representational systems and their relationship to perceived trustworthiness in a counseling analogue. *J Counsel Psychol* 1981;**28**:305–8.
26. Feldstein JC. Counselor and client sex pairing: the effects of counseling problem and counselor sex role orientation. *J Counsel Psychol* 1982;**29**:418–20.
27. Freeman HR. Differential perceptions both between and within transcribed counselors as a function of level of facilitative conditions. *J Counsel Psychol* 1980;**27**:391–4.
28. Freeman ST, Conoley CW. Training, experience, and similarity as factors of influence in preferences of deaf students for counselors. *J Counsel Psychol* 1986;**33**:164–9.
29. Grimes WR, Murdock NL. Social influence revisited: effects of counselor influence on outcome variables. *Psychotherapy* 1989;**26**:469–74.
30. Henderson WJ, Lyddon WJ. Client gender role attitudes and perception of counselor characteristics. *J Ment Health Counsel* 1997;**19**:182–90.
31. Heppner PP, Heesacke M. Interpersonal influence process in real-life counseling: investigating client perceptions, counselor experience level, and counselor power over time. *J Counsel Psychol* 1982;**29**:215–23.
32. Heppner PP, Heesacker M. Perceived counselor characteristics, client expectations, and client satisfaction with counseling. *J Counsel Psychol* 1983;**30**:31–9.
33. Hubble MA, Noble, FC, Robinson SE. The effect of counselor touch in an initial counseling session. *J Counsel Psychol* 1981;**28**:533–5.
34. Khan JA, Cross DG. Stereotypic conceptions vs perceptions of expert counsellor behaviours. *Can Counsel* 1982;**16**:201–5.
35. Khan JA, Cross DG, Dalgleish L, Wiltshire EB. Role of task and interpersonal variables in differential perception of short-term insight-oriented and behavior therapy. *Psychol Rep* 1982;**50**:1179–86.
36. Kottke JL, MacLeod CD. Use of profanity in the counseling interview. *Psychol Rep* 1989;**65**:627–34.
37. Kratz NA, Marshall LL. First impressions: analog experiment on counselor behavior and gender. *Represent Res Soc Psychol* 1988;**18**:41–50.

continued

38. LaCrosse MB. Comparative perceptions of counselor behavior: a replication and extension. *J Counsel Psychol* 1977;**24**:464–71.
39. LaFromboise TD, Dixon DN. American Indian perception of trustworthiness in a counseling interview. *J Counsel Psychol* 1981;**28**:165–9.
40. McCarthy PR, Betz NE. Differential effects of self-disclosing versus self-involving counselor statements. *J Counsel Psychol* 1978;**25**:251–6.
41. McCarthy PR, Betz NE. Differential effects of self-disclosing versus self-involving counselor statements. In Hill CE, editor. *Helping skills: the empirical foundation*. Washington, DC: American Psychological Association; 2001. pp. 389–96.
42. McKee K, Smouse AD. Clients' perceptions of counselor expertness, attractiveness, and trustworthiness: initial impact of counselor status and weight. *J Counsel Psychol* 1983;**30**:332–8.
43. Merluzzi TV, Brischetto CS. Breach of confidentiality and perceived trustworthiness of counselors. *J Counsel Psychol* 1983;**30**:245–51.
44. Milne CR, Dowd ET. Effect of interpretation style on counselor social influence. *J Counsel Psychol* 1983;**30**:603–6.
45. Moran MR. Effects of sexual orientation similarity and counselor experience level on gay men's and lesbians' perceptions of counselors. *J Counsel Psychol* 1992;**39**:247–51.
46. Neimeyer GJ, Walling CC. Perceived social influence in mental health: the professionals' perspective. *Soc Behav Person* 1990;**18**:217–24.
47. Orcutt MA, Walsh WB. Recognition of sex-biased counseling interactions. *Prof Psychol Res Pract* 1983;**14**:462–72.
48. Paurohit N, Dowd ET, Cottingham HF. The role of verbal and nonverbal cues in the formation of first impressions of black and white counselors. *J Counsel Psychol* 1982;**29**:371–8.
49. Ruppel G, Kaul TJ. Investigation of social influence theory's conception of client resistance. *J Counsel Psychol* 1982;**29**:232–9.
50. Sandhu SDS, Reeves TG, Portes PR. Cross-cultural counseling and neurolinguistic mirroring with Native American adolescents. *J Multicult Counsel Dev* 1993;**21**:106–18.
51. Sodowsky GR, Parr G. Cultural consistency and counselor credibility. *TACD J* 1991;**19**:33–8.
52. Strohmer DC, Biggs DA. Effects of counselor disability status on disabled subjects' perceptions of counselor attractiveness and expertness. *J Counsel Psychol* 1983;**30**:202–8.
53. Subich LM. Ratings of counselor expertness, attractiveness, and trustworthiness as a function of counselor sex role and subject feminist orientation. *Sex Roles* 1984;**11**:1033–43.
54. Uhlemann MR, Lee DY, France H. Counsellor ethnic differences and perceived counselling effectiveness. *Int J Advance Counsel* 1988;**11**:247–53.
55. VandeCreek L, Angstadt L. Client preferences and anticipations about counselor self-disclosure. *J Counsel Psychol* 1985;**32**:206–14.
56. Wachowiak D, Diaz S. Influence of client characteristics on initial counselor perceptions. *J Counsel Psychol* 1987;**34**:90–2.
57. Watkins CE, Schneider LJ. Self-involving versus self-disclosing counselor statements during an initial interview. *J Counsel Dev* 1989;**67**:345–9.
58. Zamostny KP, Corrigan JD, Eggert MA. Replication and extension of social influence process in counseling: a field study. *J Counsel Psychol* 1981;**28**:481–9.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C19
Counselor Rating Form – Short Version (CRF-S)

General details

Authors	Corrigan JD, Schmidt LD
Language	English
Country of publication/development	USA
Publication date	1983
Publisher	NA

Purpose and overview

The Counselor Rating Form – Short Version (CRF-S) is a revision of the Counselor Rating Form (CRF). The CRF is also covered in this review and was designed to measure client- or observer-perceived counsellor behaviour, particularly to investigate the existence of three dimensions of perceived counsellor behaviour: expertness, attractiveness and trustworthiness

The form was revised with the intentions of improving utilisation and reliability; widening access by lowering the required reading ability and making greater use of the lower end of the seven-point scale. The CRF-S has 12 items selected from the CRF's 36

Theoretical orientation	Social psychology, specifically social influence theory
Population details	Clinical and non-clinical adults, clinical adolescents
Perspective	The client or an independent observer may complete the form
Measure used by	Researchers
Other versions	Counselor Rating Form. CRF-Quick Score (CRF-QS)[9]
Notes	The initial validation of the CRF-S followed two procedures. The first was a replication of the methodology of the original CRF, where 133 volunteer students viewed and then rated Rogers, Perls and Ellis in the film *Three approaches to psychotherapy*. In the second procedure 155 clients participating in outpatient therapy completed the form after a regular scheduled interview with community counsellors[1]

Areas of therapist–patient interaction addressed: Map

Roles: confidant; expert/authority/leader; good object

Therapist engagement: warmth; genuiness

Patient engagement: attraction

The patient-therapist interaction information is derived from the form's items

Dimensions

Expertness	Four items relating to perceived counsellor expertness, which had high factor loadings in factor analyses of the CRF
Attractiveness	Four items relating to perceived counsellor attractiveness, which had high factor loadings in factor analyses of the CRF
Trustworthiness	Four items relating to perceived counsellor trustworthiness, which had high factor loadings in factor analyses of the CRF

continued

Reliability	
The CRF has demonstrated adequate split-half reliability[1] and internal consistency[2,4,6]	
Split-half	Split-half reliabilities of 0.90, 0.91 and 0.87 have been found for expertness, attractiveness and trustworthiness, respectively[1]
Internal consistency	Alpha coefficients for expertness, attractiveness and trustworthiness, respectively, have been found to be: 0.93, 0.92 and 0.87[6] 0.87, 0.86, 0.76, (collapsed ratings of three counsellors)[2] 0.93, 0.92, 0.92, total 0.95 and averages of 0.86 (range 0.82 to 0.91), 0.87 (range 0.81 to 0.93) 0.83 (range 0.77 to 0.90), total = 0.82 (range 0.65 to 0.86)[4]
Inter-rater	No details
Test–retest	No details
Validity	
The CRF-S items have face and content validity as they were originally selected for the CRF by expert judges and met the criterion of having agreement from 75% of the judges as to their appropriate category (expertness, attractiveness or trustworthiness, see CRF summary). Concurrent validity of the CRF-S has been assessed with different methods, yielding mixed results. In one study, the CRF-S demonstrated adequate concurrent validity with the CRF and the Counselor Effectiveness Rating Scale (CERS).[6] Another study that compared the CRF-S and CRF did find significant differences between the two[2] In assessments of discriminant validity, where the form's three scales were expected to be relatively independent, intercorrelations ranged from 0.54 to 0.77.[6] As the revisions to the CRF were, in part, to make greater use of the lower end of the scales, the discriminant validity of the CRF-S was assessed by comparing scores with those from the CRF. The expected differences were not found[2] The factor structure of the CRF has been assessed with interscale correlations, confirmatory and principal components factor analyses. Interscale correlations ranged from 0.27 to 0.72.[2] Two studies used confirmatory factor analysis to test competing models. While no model had a statistical fit to the data, the two studies found a three-factor oblique structure,[1] and a two-step hierarchical model[4] had the best fit. A principal components analysis revealed a two-factor model, accounting for 73% of the total variance[6]	
Face	See Content validity
Content	The form's items were originally selected for the CRF, having met the criterion of attaining agreement from 75% of the expert judges as to whether they represented counsellor expertness, attractiveness or trustworthiness
Criterion (a) concurrent	The CRF-S dimensions were correlated with their equivalent dimensions on the CRF and the CERS with the following results ($n = 160$, $p < 0.01$): CRF-S expertness: 0.81 and 0.83 with CRF and CERS, respectively CRF-S attractiveness: 0.86 and 0.87 with CRF and CERS, respectively CRF-S trustworthiness: 0.79 and 0.86 with CRF and CERS, respectively[6] A *z*-test of independent correlations indicated that both the attractiveness/expertness and attractiveness/trustworthiness correlations were significantly lower ($p < 0.05$) on the CRF-S than on the CRF[2] CRF-S and CRF ratings of Rogers, Perls and Ellis (in the film *Three approaches to psychotherapy*) were compared with the following results ($n = 215$): Ellis' expertness ratings were significantly lower on the CRF-S (5.51) than on the CRF (6.01)[2] Ellis and Perls were rated as equally expert on the CRF, but Perls was rated as more expert than Ellis on the CRF-S ($p > 0.01$)[2] Rogers was rated as less attractive than Perls and Ellis on the CRF-S ($p > 0.05$) and as more attractive than the Perls and Ellis on the CRF ($p < 0.05$)[2]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Criterion (b) predictive	No details
Construct	No details
Convergent	No details
Discriminant	Where the three CRF-S dimensions are expected to be relatively independent, the following correlation coefficients were found: Expertness/attractiveness: 0.60 to 0.72 (median 0.68) Expertness/trustworthiness: 0.54 to 0.73 (median 0.67) Attractiveness/trustworthiness: 0.66 to 0.77 (median 0.70)[6] The CRF-S was designed, in part, to rectify the underuse of the lower end of the scales found with the CRF. To test this, CRF-S and CRF scores were compared (with differences expected), with the following results (n = 215): Counsellor × Instrument ANOVA failed to reveal a significant effect for Instrument; F tests for homogeneity of variance did not reveal significant differences between the two forms in dimension means; and 16% of item ratings on the CRF-S were below the midpoint, compared to 14% on the CRF[2]
Factor analysis	CRF-S interscale correlations ranged from 0.27 to 0.72, with a median of 0.56 (p not reported)[2] Confirmatory factor analysis tested five competing models. No model statistically fitted the data, as tested by χ^2 tests, but model 5, a three-factor oblique structure, best fitted the data. The lowest intercorrelations in this model were between expertness and attractiveness, with the lowest being between expertness and trustworthiness[1] Confirmatory factor analysis tested four models. No model statistically fitted the data, as tested by χ^2 tests, but a two-step hierarchical model was found to best fit the data[4] Principal components analysis of the CRF-S revealed two principal components. Factor 1, attractiveness–trustworthiness, had eight items loading from 0.56 to 0.89, accounting for 63% of the variance. Factor 2, expertness, had four items loading from 0.77 to 0.90 and accounted for a further 10.9% of the variance. OBLIMIN rotation of these two factors revealed an attractiveness–trustworthiness factor and an expertness factor, whose correlation was r = 0.62. factor[6]
Responsiveness	
Discriminative (between individuals)	Rogers, Perls and Ellis, in the film *Three approaches to psychotherapy*, have been rated differently on the CRF-S. Perls was perceived as more expert than Ellis, who was perceived as more expert than Rogers (p 0.05)[2]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	12
Administration method	Rating scale
Time taken to complete	An average completion time of 5 minutes has been reported for the CRF–QS[9]
Flesch reading age	Eighth grade
Translations	No details
Access by ethnic minorities	The form has been used in studies with participants from ethnic minorities

continued

Feasibility	
Copyright	1983, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	Observer raters complete the form immediately after viewing video footage of a therapy session. Client raters complete the form immediately after the end of a session
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Likert scale. Each item is an adjective (e.g. friendly), with a seven-point scale anchored at not very and very
Normative data	No details
Notes	
The CRF-S items are listed in ref. 1 Other uses of the CRF-S in research include: A study of the effects of the gender of the counsellor and of gender role orientation on client ratings of training-in-training[7] An examination of the extent to which the level of reactance of adolescent clients served as a mediating factor for counsellor ratings and two paradoxical intervention techniques[8] Validation of a quick score CRF[9] The effects of presentation format on participants' evaluation in analogue studies[10] An examination of the effects of counselling styles [problem-solving counselling (PSC), client-centered counselling (CCC), and relationship-centred counseling (RCC)] and stages (beginning, working and ending) on perceived counsellor effectiveness were examined. This study employed Taiwanese female students[11] A study in which 172 male and 208 female undergraduates used the form to evaluate a hypothetical male/female psychologist to assess whether men and women held different stereotypes for male and female therapists[12] A vicarious participation counselling analogue to examine the effects of the racial identity attitudes (RIAs) of black male surrogate clients on their perceptions of the parallel counselling dyad[13] A study that compared the effects of three types of offered-offered metaphors, varying in levels of complexity, to facilitative responses on tests of perceived empathy, regard, expertness, attractiveness and trustworthiness[14]	
Résumé	
Strengths	It is short (12 items) and can be administered quickly (in about 5 minutes). Eighth grade reading level makes the form accessible. Correlations are consistently adequate for split-half[1] and internal consistency[2,4,6] reliabilities. Furthermore, the internal consistency assessments were independent from the authors. The CRF-S expertness dimension has also been responsive to the different styles of therapy demonstrated by Rogers, Ellis and Perls in the film *Three approaches to psychotherapy*[2] The CRF-S demonstrated adequate concurrent validity with the CRF and the Counselor Effectiveness Rating Scale (CERS)[6]
Weaknesses	The CRF-S does not make greater use of the lower end of the scales as intended.[2] The form was designed as a shorter version of the CRF, yet assessment of convergent validity revealed significant differences.[2] Interscale correlations are fairly high (from 0.54 to 0.77) for supposedly independent concepts[6]
Areas for further research	Further assessment of psychometric properties (e.g. inter-rater reliability and the relationship between the CRF-S and the CRF)

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Primary references

1. Corrigan JD, Schmidt LD. Development and validation of revisions in the Counselor Rating Form. *J Counsel Psychol* 1983;**30**:64–75.
2. Epperson DL, Pecnik JA. Counselor Rating Form – Short version: further validation and comparison to the long form. *J Counsel Psychol* 1985;**32**:143–6.
3. Ponterotto JG, Furlong MJ. Evaluating counselor effectiveness: a critical review of rating scale instruments. *J Counsel Psychol* 1985;**32**:597–616.
4. Tracey TJ, Glidden CE, Kototovic AM. Factor structure of the Counselor Rating Form – Short. *J Counsel Psychol* 1988;**35**:330–5.
5. Tryon GS. The Counselor Rating Form – Short Version: a factor analysis. *Measure Eval Counsel Dev* 1987;**20**:122–6.
6. Wilson FR, Yager GG. Concurrent and construct validity of three counselor social influence instruments. *Measure Eval Counsel Dev* 1990;**23**:52–66.

Secondary references

7. Ametrano IM, Pappas JG. Client perceptions of counselors-in-training: the effects of sex and gender role orientation. *Counsel Educ Supervis* 1996;**35**:190–203.
8. Blankenship BL, Eells GT, Gregory T, Carlozzi AF, Perry K, Barnes LB. Adolescent client perceptions and reactions to reframe and symptom prescription techniques. *J Ment Health Counsel* 1998;**20**:172–82.
9. Dorn FJ, Jereb R. Enhancing the usability of the Counselor Rating Form for researchers and practitioners. *Measure Eval Counsel Dev* 1985;**18**:12–16.
10. Johnson ME, Pierce CA, Baldwin K, Harris A, Brondmo AK. Presentation format in analogue studies: effects on participants' evaluation. *J Psychol* 1996;**130**:341–9.
11. Lin Y.-n. The effects of counseling styles and stages on perceived counselor effectiveness from Taiwanese female university clients. *Asian J Counsel* 2001;**8**:35–60.
12. Marshall LL, Kratz NZ. Preexisting differences in evaluations of counselors. *Psychol Rep* 1988;**63**:889–90.
13. Richardson TQ, Helms JE. The relationship of the racial identity attitudes of black men to perceptions of 'parallel' counseling dyads. *J Counsel Dev* 1994;**73**:172–7.
14. Suit JL, Paradise LV. Effects of metaphors and cognitive complexity on perceived counselor characteristics. *J Counsel Psychol* 1985;**32**:23–8.

C20
Counselor Evaluation Rating Scale (CERS)

General details	
Author	Myrick RD
Language	English
Country of publication/development	USA
Publication date	1971
Publisher	NA
Purpose and overview	
To enable a respondent (supervisor) to rate a counsellor's (student or otherwise) performance in counselling and supervision. The items concern the student's understanding of a counselling rationale, counselling practice with clients, and exploration of self and counselling relationships with their supervisor. Developed to help supervisors evaluate practicum students' behaviour and not as a criterion measure of counsellor effectiveness	
Theoretical orientation	Pan-theoretical
Population details	See below
Perspective	Counsellor self-report/independent rater
Measure used by	The CERS evaluates a variety of trainee behaviours, and is therefore a useful indicator of a trainee's performance during initial and closely supervised counselling situations[3]
Other versions	No details
Notes	*Practitioners:* Master's level students in counsellor education programmes at eight large state universities, 80% women, 71% between 25 and 45 years old; self-report[1] 19 graduate students enrolled in a supervised practicum in a counsellor education programme (11 women, age 22–52, mean age 29.7 years)[2] Students pursuing educational specialist or doctoral degrees in counsellor education[3] Student counsellors[4,6,8] 87 graduate counselling students[5] 131 beginning clinical trainee graduate students (mean age 28.0 years); self-report[9] 54 counsellors-in-training, aged 23–48 years, who were in supervised practical as part of a counselling programme at a major university[10] *Raters:* Variety of faculty supervisors[3] Supervisors from the University of Florida[4]
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: openness; genuineness; listening Framework: focused; rigid; collaborative/participative/involving Non-verbal communication: paralinguistics Threats to the relationship: critical; intrusive Outcomes: general satisfaction Inferred from fully listed items in ref. 4	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Scales	
The scale yields three scores: (1) counselling, (2) supervision and (3) total	
Counselling	13 items: designed to assess an individual's work in counselling, e.g. 'Tends to talk more than client during counselling'
Supervision	13 items: appraise the counsellor's work and progress in supervision
Total	When the items in the two subcategories are totalled and the final item on the CERS, 'Can be recommended for a counselling position without reservation' is included, the composite score is a measure of an individual's performance in a supervised counselling experience
Reliability	
The split-half validity, internal consistency and test–retest reliability of the CERS are adequate	
Split-half	Spearman Brown correction 0.95[4] 0.87[2]
Internal consistency	0.86. The 13 supervisory items were correlated with the 13 counselling items[4]
Inter-rater	No details
Test–retest	0.94. Minimum period of 4 weeks[4]
Validity	
Face and construct validity of the CERS have been addressed. Partial convergent validity was demonstrated for the CERS with Carkhuff's Communication of Respect in Interpersonal Processes scale. Two studies that have addressed the factor structure of the CERS have reached different conclusions as to the number of factors	
Face	Items were analysed, clarified and assessed in terms of their face validity as measures of effective behaviours in counselling and supervision[4] During the development of the scale, potential items were sent to faculty and students at the Department of Counselor Education, University of Florida. Some items were discarded and others developed according to feedback[4]
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	A significant positive relationship ($r = 0.42$, $p < 0.05$) between students' ratings on supervision (as measured by CERS) and the level of respect (as measured by Carkhuff's Communication of Respect in Interpersonal Processes scale) offered to their clients in counselling was found[2]
Discriminant	No details
Factor structure	An oblique principal axes factor analysis was conducted, and six primary factors emerged. These were labelled (1) general counselling performance (evaluations of a wide variety of counsellor trainee characteristics, such as comfort, awareness, confidence, sensitivity, mode of expression and overall counselling ability), (2) professional attitude (evaluations of trainees' outlooks on their own professional activities), (3) counselling behaviour (flexibility in orientation and verbal behaviour), (4) counselling knowledge (evaluations of counsellor trainees' abilities to discuss and use theoretical counselling information), (5) supervision attitude (evaluations of whether or not the trainee participates in supervision in an open, honest and self-aware manner); and (6) supervision behaviour (evaluations of trainee behaviours important to the formation of effective professional relationships). These factors accounted for 71% of the total variance. Inspection of the factor loadings indicates that a good approximation of simple structure was achieved since 23 of the 27 items had high loadings on one factor and relatively low loadings on the others. The total score (overall

continued

	supervised counselling effectiveness) seemed to have the greatest validity. Ben Hoff and Thomas (1992) found high correlations among several of these factors. For example, factors 1 and 2 were highly correlated ($r^2 > 0.50$) with each other as well as with all other factors: only factors 3 and 4 and factors 3 and 5 had low correlations ($r^2 < 0.3$)[3] A principal axis extraction method of factor analysis was conducted, and four factors emerged, accounting for 41% of the total variance before rotation. The correlation matrix from this factor analysis was compared with the one derived from Loesch and Rucker's factor model,[3] and large residuals were noted. The χ^2 statistic generated to test the statistical independence of the findings was very large ($\chi^2 = 572.33$, $p < 0.001$), indicating that the self-report data (this study) fit the Loesch and Rucker model very poorly. In this study's factor analysis factor 1 (purposeful counselling performance) accounted for 21.4% of the variance and reflects an evaluation of the overall counselling effectiveness of the training-in-training, and includes considerations such as trainee's comfort level, ability to address both content and feeling, flexibility and spontaneity, understanding of the counselling process, performance in supervision, and level of self-confidence. This factor 1 is similar to Loesch and Rucker's factor 1. The second factor (non-counselling behaviours) describes characteristics that are undesirable for a professional counsellor, e.g. lack of sensitivity to dynamics of self in supervisory relationships, or mechanical, rigid counselling behaviour. Three of the seven items of this factor constitute Loesch and Rucker's counselling behaviour (factor 3). The third factor (supervision attitude) is related only in name to Loesch and Rucker's fifth factor, as only one item appears in both. Factor 4 (counselling orientation) is composed of three items that describe the counsellor's ability to consult with supervisors and colleagues when necessary, as well as the ability to keep the focus on the client during a counselling session. These items appear separately in factors 2, 6 and 3 in the Loesch and Rucker analysis. The current analysis resulted in four relatively independent factors that, except for factor 3, tend to blend together counselling and supervision attitudes and behaviours. By contrast, Loesch and Rucker (1977) found six factors that tended to be much more clearly related to either counselling or supervision subscales[1]
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	27
Administration method	Questionnaire. Self-administration of the CERS is desirable because it provides a way for counsellors-in-training formally to assess their own skills and development over the course of their graduate experiences[1]
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Feasibility	
Copyright	1971, American Counseling Association
Web or scanning options	No details
Training details	The CERS is relatively easy to fill out[4]
Administration/process details	The evaluation of effectiveness in counselling is obtained by adding items marked with a 'C', and similarly an 'S' for supervision. The range scores for each category (counselling and supervision) are 13–91 and for the total 27–189. Space is also provided on the instrument for additional comments or elaboration Participants rated themselves typically during the middle third of the semester, using optically scannable answer forms, which were scored using a computer program based on instructions provided by the test authors[1]
Support from measure developers	Questionnaires can be scored using a computer program based on instructions provided by the test authors
FAQ facility	No details
Precision	
Scale type	Ordinal. Likert. The ratings range from +3 (strongly agree) to –3 (strongly disagree). For scoring purposes, a –3 is given a scaled score of 1, while +3 is given a scaled score of 7. A reply of 'uncertain', 'not able to judge', or no response is given a median score of 4. Nine items were randomly selected and negatively expressed in an attempt to counter the effects of a possible response set
Normative data	No details
Résumé	
Strengths	Provides feedback for trainees on their performance in both counselling and supervision[3,4] The CERS offers a relatively standardised approach for conceptualising the counsellor's performance and progress[4] The CERS can be used to stimulate discussion and communication between supervisor and counsellor[4]
Weaknesses	The CERS does not achieve all of its authors' primary purposes. Six primary factors emerged from the factor analysis, instead of the two that would be expected of the items related only to the student's total counselling or supervision performances. Users should be advised to proceed cautiously when making conclusions based on the counselling and supervision subscales[3] The results of a confirmatory factor analysis call into question the claims made for the CERS by its authors that the instrument is useful for self-ratings by counsellors-in-training. The findings suggest that when counsellors use the CERS to rate themselves, different factors may emerge than when experienced supervisors use this instrument to evaluate supervisee progress and performance. CERS users should be cautious about interpreting results in terms of the Loesch and Rucker (1977) factors when CERS respondents are rating themselves. The CERS may be primarily measuring overall counselling performance, at least when counsellors-in-training rate themselves, as opposed to the subscales of counselling and supervision[1]
Areas for further research	Lack of research on the CERS when self-administered.[1] Whether the CERS, when used as a self-rating instrument, is more useful for post-master's students than for master's students, because of the post-master's students' greater experience with counselling and supervision[1]

continued

Primary references

1. Benshoff JM, Thomas WP. A new look at the Counselor Evaluation Rating Scale. *Counsel Educ Supervis* 1992;**32**:12–22.
2. Jones LK. The Counselor Evaluation Rating Scale: a valid criterion of counselor effectiveness? *Counsel Educ Supervis* 1974;**14**:112–16.
3. Loesch LC, Rucker BB. A factor analysis of the Counselor Evaluation Rating Scale. *Counsel Educ Supervis* 1977;**16**:209–16.
4. Myrick RD, Kelly FD Jr. A scale for evaluating practicum students in counseling and supervision. *Counsel Educ Supervis* 1971;**10**:330–6.

Secondary references

5. Benshoff JM. Peer supervision in counselor training. *Clin Supervis* 1993;**11**:89–102.
6. Borders LD, Fong ML. Evaluations of supervisees: brief commentary and research report. *Clin Supervis* 1991;**9**:43–51.
7. Dodenhoff JT. Interpersonal attraction and direct–indirect supervisor influence as predictors of counselor trainee effectiveness. *J Counsel Psychol* 1981;**28**:47–52.
8. Manthei RJ. The response-shift bias in a counsellor education programme. *Br J Guid Counsel* 1997;**25**:229–37.
9. Meier ST. Investigating clinical trainee development through item analysis of self-reported skills: the identification of perceived credibility. *Clin Supervis* 2001;**20**:25–37.
10. Watts RE, Trusty J. Social interest and counselor effectiveness: an exploratory study. *Indiv Psychol* 1995;**51**:293–8.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

C21
Counselor Perception Questionnaire (CPQ)

General details	
Author	Blocher D
Language	English
Country of publication/development	USA
Publication date	1985
Publisher	NA
Purpose and overview	
To measure outcomes, in terms of counsellor cognitive growth, of developmental programmes of counseling supervision	
Theoretical orientation	Cognitive-developmental[1]
Population details	See below
Perspective	Counsellors' perceptions of client behaviour. These are then rated by independent judges
Measure used by	Practitioners
Other versions	No details
Notes	*Client:* Young woman approaching college graduation, indecisive about two job offers *Practitioners:* Masters students and experienced counselling psychologists
Areas of therapist–patient interaction addressed: Map	
Non-verbal communication: paralinguistics	
Emotional expression: expression of feelings	
Inferred from items	
Dimensions	
Differentiation (D)	Yields two independent scores
Integration (I)	
Reliability	
Inter-rater reliability scores were adequate/high. No other areas of reliability were addressed	
Split-half	No details
Internal consistency	The CPQ I and D scores are moderately related to each other. They are sufficiently different, however, to merit separate use (no figures provided)
Inter-rater	Product moment correlations. Two judges rated ten master's level counselling students. Reliabilities were 0.97 and 0.94 for D and I scores, respectively
Test–retest	No details

continued

Validity	
Both the CPQ I and D scores are essentially unrelated to Paragraph Completion Method (PCM) scores, demonstrating discriminant validity. Both CPQ scores have substantial correlations with the Crockett score, reflecting the number of constructs used in a person perception task and demonstrating adequate convergent validity	
Face	No details
Content	60 case summaries were analysed following Crockett *et al*. (1973). Scoring methods were devised to incorporate measures of both cognitive complexity and veridicality directly in content that was relevant to issues in counselling supervision
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	The criterion variable for convergent validity was the Crocket *et al*. (1973) complexity measure. Correlations between Crocket and CPQ I and D were 0.78 and 0.56, respectively
Discriminant	Correlations were calculated of 14 master's students' scores on the CPQ and the PCM. The CPQ is intended to be a measure of complexity of person perception and should be unrelated to PCM scores, which measure general conceptual levels. Correlations were 0.01 and –0.01 between PCM and CPQ I and D scores respectively
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	CPQ mean scores of eight experienced PhD counselling psychologists were compared (*t*-test) with those of ten first year counselling students. The PhDs scored higher on both the D and I scores, but only the D score was statistically significant with the small samples
Evaluative (within individual across time)	No details
Acceptability	
Number of items	NA
Administration method	Based on responses to videotaped counselling interview material
Time taken to complete	50 minutes
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	No details
Web or scanning options	No details
Training details	Scoring directions provided
Administration/process details	Participants see three 5-minute excerpts from a vocational counselling interview, with 5 minutes after each excerpt to respond. At the end of the series of three segments, participants have 20 minutes to respond to the total interview Independent raters then score these responses in terms of cues, e.g. gestures or hands
Support from measure developers	No details
FAQ facility	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Precision	
Scale type	Rating scale. Nominal, binary
Normative data	No details
Résumé	
Strengths	The single study on the CPQ demonstrates good responsiveness, adequate convergent and discriminant validity, and adequate inter-rater reliability. Extensive training is not required
Weaknesses	Administration of the questionnaire takes 50 minutes and then it needs to be rated
Areas for further research	Much more work needs to be done in refining and validating this instrument[1]
Primary reference	
1. Blocher D, Christensen EW, Hale-Fiske R, Neren SH, Spencer T, Fowlkes S. Development and preliminary validation of an instrument to measure cognitive growth. *Counsel Educ Supervis* 1985;**25**:21–30.	
Secondary references	
None	

C22
Cross-Cultural Counseling Inventory – Revised (CCCI-R)

General details	
Author	LaFromboise TD
Language	English
Country of publication/development	USA
Publication date	1991
Publisher	NA
Purpose and overview	
A measure of cross-cultural competence, developed to meet the need for explicit assessment of counselling effectiveness with culturally diverse clients. The CCCI-R was devised to respond to a perceived need in the field for an instrument capable of both assessing a counsellor's ability to deal effectively with clients from diverse ethnic and cultural groups, and evaluating the efficacy of cross-cultural counselling and training models and methods	
Theoretical orientation	Counselling
Population details	Counsellors/trainee counsellors
Perspective	Independent rater
Measure used by	Trained counsellors
Other versions	Original Cross-Cultural Counseling Inventory 22 items and a short 12-item version
Notes	
Areas of therapist–patient interaction addressed: Map	
Broader context: diversity; religion	
Therapy context: values; responsibilities	
Roles: advocate; protector	
Individual differences: problem complexity; social support	
Therapist engagement: respect; support/tolerance	
Framework: structuring; directive	
Outcomes: achieving a working relationship	
Information derived from items	
Dimensions	
Awareness and beliefs	Six items. Counsellor's sensitivity to his or her personal values and biases and how these may influence perceptions of the client
Knowledge	Four items. Counsellor's knowledge of the client's culture, worldview and expectations for the counselling relationship
Skills	Ten items. Counsellor's ability to intervene in a manner that is culturally sensitive to others
Reliability	
The CCCI-R demonstrates adequate internal consistency and partial to adequate inter-rater reliability. Inter-rater reliability is higher when expert/highly experienced raters are used	
Split-half	No details
Internal consistency	Coefficient alpha (0.95) (ref. 2, study 3[1,3,4])

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	Coefficient alpha levels were 0.88, 0.92 and 0.95 across undergraduate students, graduate students and faculty[3,5]
Inter-rater	Kappa was 0.58 ($p < 0.001$) among eight judges (ref. 2, study 1[3])
	Correlations among three raters ranged from 0.39 to 0.69; average rating across three raters was 0.78 (ref. 2, study 2)
	Correlations among three expert judges rating 12 videotaped counselling vignettes was 0.78, rising to 0.84 when one of the videotapes in which there was particularly poor agreement was discarded[4,5]
	Using the Spearman–Brown prophesy formula, estimated reliability for a single rater is 0.54, rising to 0.63 with the removal of a low-quality tape[5]
Test–retest	No details
Validity	
The CCCI-R has evidence for face and content validity. The instrument has demonstrated adequate discriminant validity with the Counselor Rating Form (CRF). Evidence for the factor structure is mixed	
Face	CCCI-R is based on the 11 cross-cultural counselling competencies outlined in a position paper by the Education and Training Committee of the Division of Counseling Psychology of the American Psychological Association
Content	The overall level of agreement of eight raters classifying each CCCI-R item in accordance with its original intent, as defined by Division 17 11 cross-cultural counselling competencies, was 80%. This level of agreement demonstrates that the CCCI-R has acceptable content validity and is representative of the domain of cross-cultural counselling competence (ref. 2, study 1[3])
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	Raters ($n = 86$) rated as above average a counsellor judged by her faculty to have a high level of cross-cultural counselling competence[2]
	Students who are perceived as more culturally competent by their clinical supervisors score higher on the scale when scored by independent judges[3]
Convergent	No details
Discriminant	The original CCCI scores correlated minimally (0.01 to 0.28) with the CRF (Barak and LaCrosse, 1975), which is a general measure of a counsellor's expertness, trustworthiness and attractiveness[3–5]
Factor structure	With a sample of 86 counselling students LaFromboise *et al*.[2] examined the factor structure of the CCCI-R using a principal components technique with squared multiple correlations as initial communality estimates. An orthoganol rotation indicated three factors with eigenvalues higher than 1.0. A scree test subsequently indicated a single factor accounting for 51% of the scale variance. 19 of the 20 scale items loaded (0.55 or above factor loading) on this factor. A second factor analysis was conducted in an attempt to isolate distinctive features of the CCCI-R. This analysis resulted in a three-factor solution accounting for 63% of the variance. The three factors emerging were labelled cross-cultural counselling skill, sociopolitical awareness, and cultural sensitivity.[3–5] See ref. 2 for full details of the factor analysis
Responsiveness	
Discriminative (between individuals)	Raters ($n = 86$) rated as above average a counsellor judged by her faculty to have a high level of cross-cultural counselling competence[2]
	Students who are perceived as more culturally competent by their clinical supervisors score higher on the scale when scored by independent judges[3]
Evaluative (within individual across time)	No details

continued

Acceptability	
Number of items	20
Administration method	Questionnaire
Time taken to complete	25 minutes
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	Used with all ethnic groups
Feasibility	
Copyright	1991, American Psychological Association
Web or scanning options	No details
Training details	No training provided
Administration/process details	Videotaped segments are viewed and rated by trained counsellors. Time of segments ranges from 7 to 15–20 minutes
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Six-point Likert type scale ranging from 1 (strongly disagree) to 6 (strongly agree). Higher scores indicate greater ability to work with clients from diverse racial/ethnic groups
Normative data	No details
Résumé	
Strengths	Adequate levels of internal consistency. Evidence for content and construct validity/responsiveness. Brevity of the scale makes it easy and efficient to score
Weaknesses	Inter-rater reliability is dependent on the expertise of the raters. Mixed evidence for factor structure. Measure developers recommend the scale be scored as a unidimensional construct
Areas for further research	Additional factor analytic studies of the CCCI-R are needed using large, geographically dispersed samples
	Research required on test–retest reliability

Primary references

1. Boyle DP, Springer A. Toward a cultural competence measure for social work with specific populations. *J Ethn Cult Divers Soc Work* 2001;**9**(3/4):53–71.
2. LaFromboise TD, Coleman HL, Hernandez A. Development and factor structure of the Cross-Cultural Counseling Inventory – Revised. *Prof Psychol Res Pract* 1991;**22**:380–8.
3. Ponterotto JG, Casas JM. *Handbook of racial/ethnic minority counseling research*. Springfield, IL: CC Thomas; 1991.
4. Ponterotto JG, Rieger BP, Barrett A, Sparks R. Assessing multicultural counseling competence: a review of instrumentation. *J Counsel Dev* 1994;**72**:316–22.
5. Sabnani HB, Ponterotto JG. Racial/ethnic minority-specific instrumentation in counseling research: a review, critique, and recommendations. *Measure Eval Counsel Dev* 1992;**24**:161–87.

Secondary references

None

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

E1
Empathy Construct Rating Scale (ECRS) – 23 items

General details	
Authors	Hughes R, Hukill R
Language	English
Country of publication/development	USA
Publication date	1982
Publisher	NA

Purpose and overview	
A questionnaire rating scale designed to measure empathy in nursing contexts	
Theoretical orientation	Mental health nursing
Population details	Adult nursing client
Perspective	Client, nurse and peer report
Measure used by	Nursing professionals
Other versions	84-item version
Notes	

Areas of therapist–patient interaction addressed: Map

Therapy context: boundaries; values; responsibilities

Therapist engagement: empathy/sensitivity; warmth; genuineness; respect; support/tolerance; openness; listening

Framework: convergent; complementary; reciprocal; congruent; controlling; flexible/rigid

Therapeutic techniques: responsiveness/receptiveness/attunement

Threats to the relationship: defensive; critical; hostility/anger

Inferred from full list of items

Dimensions

None specified

Reliability	
As measured by Cronbach's alpha estimate, the ECRS demonstrated adequate internal reliability[1,2]	
Split-half	No details
Internal consistency	$\alpha = 0.84$,[1] 0.88[2]
Inter-rater	NA
Test–retest	No details

Validity

Concurrent validity of the ECRS was tested with three empathy instruments: the Carkhuff Empathic Understanding Scale, the empathy subtest of the Barrett-Lennard Relationship Inventory (BLRI) and the Empathy Test (Layton, 1979). The ECRS displayed adequate convergent validity with the empathy test of the BLRI and partial convergent validity with the Carkhuff scale[1]

Construct validity was demonstrated by finding support of the hypothesis that registered nurses (RNs) would perform better on the ECRS than nursing assistants (NAs) owing to greater experience and training[1]

continued

Face	No details
Content	No details
Criterion (a) concurrent	The ECRS correlated significantly with the Carkhuff scale ($r = 0.37$, $p < 0.01$) and the empathy subtest of the BLRI ($r = 0.78$, $p < 0.001$). The ECRS correlated non-significantly with the Empathy Test ($r = 0.03$)
Criterion (b) predictive	No details
Construct	RNs scored significantly better on the ECRS than NAs (RN average score 112.83, SD 16.46; NA average score 100.91, SD 23.10)
Convergent	See Concurrent
Discriminant	No details
Factor structure	No factors
Responsiveness	
Discriminative (between individuals)	The ECRS discriminated between RNs and NAs. RNs displayed significantly more empathy on the ECRS than NAs[1]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	23
Administration method	Self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1982, University of Texas
Web or scanning options	No details
Training details	No details
Administration/process details	Paper and pencil self-administered questionnaire. The respondents read each statement and decide the degree to which the statement is like/unlike their perceptions of themselves, their nurse or their peer
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	A six-point Likert scale was applied to each of the items (+3 = extremely like to –3 = extremely unlike). Negative scores are reversed and then all of the item scores are added to yield an overall empathy score, with higher scores denoting well-developed empathy
Normative data	No details
Résumé	
Strengths	Adequate reliability. Strong convergent validity with the BLRI
Weaknesses	Instrument has not been used in a wide variety of settings; is restricted to nursing studies
Areas for further research	Further testing of psychometric properties. Use of the instrument to measure change over time in levels of empathy

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Primary reference

1. Layton JM, Wykle MH. A validity study of four empathy instruments. *Res Nurs Health* 1990;**13**:319–25.

Secondary reference

2. Hughes R, Hukill R. *Participant characteristics, change and outcomes in pre-service clinical teacher education*. ERIC Document Reproduction Service No. ED 240 096. Research for Teacher Education, University of Texas at Austin; 1982.

E2
Empathy Construct Rating Scale (ECRS) – 84 items

General details	
Author	La Monica E
Language	English
Country of publication/development	USA
Publication date	1981
Publisher	NA
Purpose and overview	
The ECRS was designed to measure empathy. The items deal with a person's feelings or actions towards another person	
Theoretical orientation	Mental health nursing
Population detail	Adult clients
Perspective	Client, nurse and peer report
Measure used by	Nursing professionals
Other versions	23-item version
Notes	
Areas of therapist–patient interaction addressed: Map	
Therapy context: boundaries; values; responsibilities	
Therapist engagement: empathy/sensitivity; warmth; genuineness; respect; support/tolerance; openness; listening	
Framework: convergent; complementary; reciprocal; congruent; controlling; flexible/rigid	
Therapeutic techniques: responsiveness/receptiveness/attunement	
Threats to the relationship: defensive; critical; hostility/anger	
Inferred from general description of scale	
Dimensions	
Well-developed empathy	49 items. Positive empathy items
Lack of empathy	35 items. Negative empathy items
Reliability	
Split-half reliability estimates were produced for Form A (54 positively worded items) and Form B (46 negatively worded items). The split-half method corrected by the Spearman–Brown formula was used. High reliability coefficients resulted for both forms and on the basis of this result, all 100 items were left in the ECRS. The 100-item version, was reduced to an 84-item version, referred to below	
The internal consistencies of the total scale (84 items) and two main factors were estimated using Cronbach's coefficient alpha. Internal consistencies were computed for self-report, peer ratings, client ratings and combined ratings. All internal consistencies were adequate[1]	
Split-half	Form A = 0.89 Form B = 0.96
Internal consistency	Well-developed empathy (49 items): average 0.96; range 0.95 to 0.97
	Lack of empathy (35 items): average 0.93; range 0.90 to 0.95
	Total scale: (84 items): average 0.97; range 0.96 to 0.98
Inter-rater	NA
Test–retest	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Validity	
Concurrent validity of the ECRS was assessed with the following instruments: Carkhuff's Index of Communication (CIC); California Psychological Inventory (CPI); Human-Heartedness Questionnaire HHQ); Chapin Social Insight Test (CSI); Philosophy of Human Nature (PHN); Vocabulary Test-GT; Tennessee Self-Concept Scale. No concurrent validity was demonstrated	
Face	No details
Content	A rigorous rating process by nurse and psychology graduates and experts was undertaken to generate items and ensure the items had a good content validity. The items were then organised into five subscales of (1) non-verbal behavior, (2) personality traits, (3) sensitivity, (4) responding, and (5) respect for self and others
Criterion (a) concurrent	None demonstrated
Criterion (b) predictive	No details
Construct	No details
Convergent	No details
Discriminant	Discriminant validity between empathy as rated by self and empathy as rated by client
	Low correlations of ECRS with personality traits
Factor structure	The validity of the five subscales was not validated. Two factors of well-developed empathy and lack of empathy had high internal consistencies and accounted for 69.8 and 13.3 % of the variance, respectively
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	84
Administration method	Self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1981, *Research in Nursing and Health,* Wiley
Web or scanning options	No details
Training details	No details
Administration/process details	Paper and pencil self-administered questionnaire. The respondents read each statement and decide the degree to which the statement is like/unlike their perceptions of themselves, their nurse or their peer
Support from measure developers	No details
FAQ facility	No details

continued

Precision	
Scale type	A six-point Likert scale was applied to each of the 84 items (+3 = extremely like to –3 = extremely unlike). Negative scores are reversed and then all of the item scores are added to yield an overall empathy score, with higher scores denoting well-developed empathy
Normative data	No details
Résumé	
Strengths	Adequate internal consistency. Proven content validity with rating process. Discriminant validity with personality trait measures
Weaknesses	Length of instrument (84 items). ECRS does not examine interactions/transactions so cannot get at process[2]
	Scoring system has no 'does not apply' option, meaning that nurse and patient may be judging a perception of what the nurse is generally like
	Inadequate concurrent validity
Areas for further research	Further validity research with multiple empathy instruments. Use of the ECRS in longitudinal research
Primary reference	
1. LaMonica E. Construct validity of an empathy instrument. *Res Nurs Health* 1981;**4**:389–400.	
Secondary reference	
2. Bennett JA. 'Methodological notes on empathy': further considerations. *ANS Adv Nurs Sci* 1995;**18**:36–50.	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

E3
Empathy Test (ET)

General details	
Author	Layton JM
Language	English
Country of publication/development	USA
Publication date	1979
Publisher	NA
Purpose and overview	
A test of knowledge of principles of empathy originally designed as part of a research project that used modelling to teach empathy to nursing students (Layton, 1979).[1] Developed to determine whether knowledge or rules of empathy are learned through observing models	
Theoretical orientation	Based on Rogerian/person-centred definition of empathy
Population details	Simulated clients (adults) were used[1]
Perspective	Self-report (nurse)
Measure used by	Psychiatric nurses completed the questionnaire as a self-report. Sample in ref. 1 was 18 registered nurses (RNs) and 32 nursing assistants (NAs). Most were female
Other versions	No details
Notes	No details
Areas of therapist–patient Interaction addressed: Map	
Therapist engagement: empathy/sensitivity; listening (developing the relationship)	
Therapeutic techniques: responsiveness/receptiveness/attunement; reflection in action; feedback; exploration (maintaining the relationship)	
Inferred from general description of the scale	
Dimensions	
NA	
Reliability	
There is limited information on the reliability of this measure	
Split-half	0.68, as reported in Layton (1979)
Internal consistency	No details
Inter-rater	NA
Test–retest	No details
Validity	
The limited information available on this measure suggested that face and content validity have been addressed and that the measure discriminates between trained and novice nurses (see Responsiveness)	
Face	Items based on Rogers' (1957, 1975) definition of empathy
Content	The fit of the items with Rogers' definition of empathy was judged to be adequate by two experts
Criterion (a) concurrent	No details
Criterion (b) predictive	No details

continued

Construct	See Convergent validity and Responsiveness
Convergent	The Carkhuff Empathic Understanding scale was significantly correlated with the Empathy Test ($r = 0.25$, $p < 0.05$). The Empathy Test did not correlate significantly with the empathy subtest of the Barrett–Lennard Relationship Inventory ($r = 0.04$) or the Empathy Construct Scale ($r = 0.03$)
Discriminant	Average discrimination of items using upper and lower groups is 33% ('reasonably good')
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	The Empathy Test discriminated between RNs and NAs: RNs scored significantly better. It was hypothesised that RNs would score significantly higher than NAs
Evaluative (within individual across time)	No details
Acceptability	
Number of items	17
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	*Research in Nursing and Health* (Layton, 1979)[2,3]
Web or scanning options	No details
Training details	No details
Administration/process details	Paper and pencil questionnaire
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	No details on present version
Normative data	No details
Notes	
The original version had two equivalent forms, each with 24 items: 12 true/false and 12 multiple choice. The test has been refined over five studies to improve discrimination and reliability. The present/second revision is a single test with 17 items All information in this summary is taken from ref. 1. The modelling study is detailed further in ref. 2. The original version is reproduced in ref. 3	
Résumé	
Strengths	The measure is relevant for nursing training and research. It is relatively brief
Weaknesses	There is little literature on this measure as applied specifically in the field of mental health, and a lack of recent literature was identified
Areas for further research	Further validation work needed. At the time of publication of ref. 1, more data were being collected and development was continuing

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Primary reference

1. Layton JM, Wykle MH. A validity study of four empathy instruments. *Res Nurs Health* 1990;**13**:319–25.

Secondary references

2. Layton JM, The use of modelling to teach empathy to nursing students. *Res Nurs Health* 1979;**2**:163–76.
3. Ward MJ, Fetler ME. *Instruments for use in nursing education research*. Boulder, CO: Western Interstate Commission for Higher Education; 1979.

E4
Experiencing Scale (EXP)

General details	
Authors	Klein MH, Mathieu-Couglan P, Kiesler DJ
Language	English
Country of publication/development	USA
Publication date	1970
Publisher	NA
Purpose and overview	
To capture the essential quality of a client's involvement in psychotherapy. The concept of 'experiencing' refers to the quality of a person's participation in therapy: the extent to which inner referents become the felt data of attention, and the degree to which efforts are made to focus on, expand and probe those data. The scale attempts to measure the way that these theoretically important levels of experiencing appear and are referred to in the client's speech during the therapy sessions. It is the patient's verbal behaviour exclusively that is rated	
The Patient Experiencing Scale (EXP) consists of one seven-point scale designed to be applied to tape-recordings or transcripts of psychotherapy. The seven scale 'stages' define the progression of client involvement in inner referents from (1) impersonal or (2) superficial, through (3) externalised or limited references to feelings, to (4) direct inner referents, to (5) questioning an unclear inner referent, to (6) focusing with a step of resolution, and finally to (7) the point where focusing comes easily and provides the connection for inner discourse	
Theoretical orientation	Pan-theoretical
Population details	Adults, couples,[7] hospitalised schizophrenics and psychoneurotic outpatients,[16,17] group psychotherapy patients[19,22]
Perspective	Observer rated
Measure used by	Counsellors; psychotherapists; research therapists. Also applied to other interactional formats (monologues/interviews) and written materials (e.g. personal documents)
Other versions	Revised version 1983
Notes	
Areas of therapist–patient interaction addressed: Map	
Patient engagement: motivation, commitment, intentions	
Threats to the relationship: defensive; resistance; withdrawal	
Outcomes: changing view of self with others	
Inferred from general description of the scale	
Dimensions	
None	
Reliability	
Adequate inter-rater reliability has been established (after training) across a range of studies, irrespective of the experience of judges	
Split-half	NA
Internal consistency	NA
Inter-rater	Reliability coefficients (*rkks*) ranged from 0.75 to 0.97 for undergraduate raters and from 0.88 to 0.99 for professional raters after training[2]
	Inter-rater reliabilities for experienced and inexperienced judges ranged from 0.91 to 0.94[1]
	A senior author and co-rater attained correlations of 0.96 and 0.89 with expert raters, respectively[3]
Test–retest	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Validity	
Several aspects of the validity of the EXP have been established across a number of studies, as detailed below	
Face	No details
Content	The experiencing scale was developed from Rogers' original client-centred theories to capture the essential quality of a client's involvement in psychotherapy. The authors have tried to keep the scale definitions and rating instructions free of diagnostic details, presenting complaints/problems, personality, specific affective state[2]
Criterion (a) concurrent	The original purpose for which the scale was developed was to test the relationship of EXP to the three therapist 'conditions' variables: positive regard, empathy and congruence, as defined by Rogers (1957). In the Wisconsin Project (Rogers *et al*., 1967) the most consistent relationships were found between EXP and Accurate Empathy (Truax and Carkhuff, 1967) rated from tapes and congruence as perceived by the patient on the Barrett-Lennard Relationship Inventory. Since the Wisconsin Project, as research proliferated, results have become more complex and mixed[2]
Criterion (b) predictive	Association of experiencing to therapeutic outcome has been shown for EXP levels at various points in therapy, most consistently at points after the first few sessions[2]
Construct	Higher levels of experiencing were found in conjunction with 'helpful' or dynamically apt therapist interventions in different kinds of individual therapy and with explicit experiential exercise in Gestalt therapy[2]
	It was hypothesised that clinical experience would be irrelevant to the EXP rating task as the authors had aimed to keep the level of clinical inference in the scale to a minimum. It was found that there were no differences between the ratings of experienced and inexperienced judges after training[1]
Convergent	The Free Association scale was correlated 0.45 with the EXP. The individual scales (involvement, freedom and spontaneity) were correlated with the EXP (range 0.31 to 0.54). The spontaneity scale was the only scale to not be significantly correlated at the 0.01 level[3]
	The EXP has been closely related to neuroticism, introspectiveness and cognitive complexity, suggesting that the scale is a measure of reflective or self-observational style[2]
Discriminant	The EXP has been shown to be weakly related to affective distress, suggesting that the scale is not a measure of expressiveness[2]
Factor structure	NA
Responsiveness	
Discriminative (between individuals)	EXP tends to be associated with neuroticism, introspectiveness, obsessiveness and self-consciousness in both help-seeking and non-help-seeking samples. EXP has also been associated with measures of cognitive style – complexity and differentiation – as well as with other indicators of reflectiveness, expressive capacity of attraction to psychotherapy[2]
	Higher levels of patient experiencing have been associated with high facilitative therapists[2]
Evaluative (within individual across time)	Improvement over therapy in levels of experiencing was demonstrated in two counselling projects[2]

continued

Acceptability	
Number of items	The EXP consists of one seven-point scale
Administration method	Judge-rated scale
Time taken to complete	No details
Flesch reading age	It is important for the raters to have good language skills since the EXP is an assessment of verbal expression
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1979, University of Wisconsin
Web or scanning options	No details
Training details	*The Experiencing Scale: a research and training manual* contains explicit procedures and materials for rater training. The formal training programme for raters is divided into eight 2-hour sessions
Administration/process details	Raters should be thoroughly trained to achieve an acceptable level of inter-rater reliability before embarking on any data collection task
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	The Experiencing Scale consists of one seven-point ordinal scale. When the rater determines that a change in EXP is taking place a different scale level is assigned
Normative data	See ref. 2
Notes	
The Experiencing Scale has often been used alongside measures of non-verbal behaviour[5,6]	
The Experiencing Scale is pan-theoretical, but has been used frequently in the study of Gestalt therapy, specifically the two-chair dialogue[10–14]	
Résumé	
Strengths	Adequate internal consistency and adequate inter-rater reliability after training. Extensive demonstration of the instrument's validity
Weaknesses	Length of training required
Areas for further research	Development and validation of the Therapist Experiencing scale
Primary references	

1. Kiesler DJ. Comparison of Experiencing Scale ratings of naive versus clinically sophisticated judges. *J Consult Clin Psychol* 1970;**35**(1).
2. Klein MH, Mathieu-Coughlan P, Kiesler DJ. The experiencing scales. In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford Press; 1986. pp. 21–71.
3. Lansford E, Bordin ES. A research note on the relation between the Free Association and Experiencing Scales. *J Consult Clin Psychol* 1983;**51**:367–9.
4. Klein MH, Mathieu PL, Gendlin E, Kiesler DJ. *The Experiencing Scale: a research and training manual*. Madison, WI: Wisconsin Psychiatric Institute; 1970.

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Secondary references

5. Davis M, Hadiks D. Nonverbal behavior and client state changes during psychotherapy. *J Clin Psychol* 1990;**46**;340–51.
6. Davis M, Hadiks D. Nonverbal aspects of therapist attunement. *J Clin Psychol* 1994;**50**:393–405.
7. De Chenne TK. Experiential facilitation in conjoint marriage counseling. *Psychother Theory Res Pract* 1973;**10**:212–14.
8. Gazzola N, Stalikas A. An investigation of counselor interpretations in client-centered therapy. *J Psychother Integr* 1997;**7**:313–27.
9. Gilliland BE. Small group counseling with Negro adolescents in a public high school. *J Counsel Psychol* 1968;**15**:147–52.
10. Greenberg LS. The intensive analysis of recurring events from the practice of Gestalt therapy. *Psychother Theory Res Pract* 1980;**17**:143–52.
11. Greenberg LS, Clarke KM. Differential effects of the two-chair experiment and empathic reflections at a conflict marker. *J Counsel Psychol* 1979;**26**:1–8.
12. Greenberg LS, Foerster FS. Task analysis exemplified: the process of resolving unfinished business. *J Consult Clin Psychol* 1996;**64**:439–46.
13. Greenberg LS, Higgins HM. Effects of two-chair dialogue and focusing on conflict resolution. *J Counsel Psychol* 1980;**27**:221–4.
14. Greenberg LS. Toward a task analysis of conflict resolution in Gestalt therapy. *Psychother Theory Res Pract* 1983;**20**:190–201.
15. Gross WF, Deridder LM. Significant movement in comparatively short-term counseling. *J Counsel Psychol* 1966;**13**:98–9.
16. Kiesler DJ. Patient experiencing and successful outcome in individual psychotherapy of schizophrenics and psychoneurotics. *J Consult Clin Psychol* 1971;**37**:370–85.
17. Levitt H, Angus L. Psychotherapy process measure research and the evaluation of psychotherapy orientation: a narrative analysis. *J Psychother Integr* 1999;**9**:279–300.
18. Levitt H, Korman Y, Angus L. A metaphor analysis in treatments of depression: metaphor as a marker of change. *Counsel Psychol Q* 2000;**13**:23–35.
19. Lewis CM, Beck AP. Experiencing level in the process of group development. *Group* 1983;**7**:18–26.
20. Mahrer AR, Stalikas A, Boissoneault M, Trainor K, Pilloud L. A scale for assessing degree of strength of client feeling. *Can J Counsel* 1990;**24**:107–16.
21. Muran JC, Samstag LW, Ventur ED, Segal ZV, Winston A. A cognitive-interpersonal case study of a self. *J Clin Psychol* 2001;**57**:307–30.
22. Nichols MP. The delayed impact of group therapists' interventions. *J Clin Psychol* 1977;**33**:258–62.
23. Schaeffer ND, Abeles N. Client attraction and distress: unexpected impact on psychotherapeutic process. *Psychother Theory Res Pract* 1977;**14**:134–8.
24. Schoeninger DW, Klein MH, Mathieu PL. Sampling from recorded therapy interview: patient experiencing ratings made with and without therapist speech cues. *Psychol Rep* 1967;**20**(1).
25. Sells DJ, Martin RB. Gender and modality differences in experiencing and emotional expression. *Can J Counsel* 2001;**35**:176–88.
26. Sherman E, Skinner KW. Client language and clinical process: a cognitive-semantic analysis. *Clin Soc Work J* 1988;**16**:391–405.
27. Slack WV, Slack CW. Talking to a computer about emotional problems: a comparative study. *Psychother Theory Res Pract* 1977;**14**:156–64.
28. Stalikas A, Fitzpatrick M. Client good moments: an intensive analysis of a single session. *Can J Counsel* 1995;**29**:160–75.
29. Stalikas A, Fitzpatrick M. Relationships between counsellor interventions, client experiencing, and emotional expressiveness: an exploratory study. *Can J Counsel* 1996;**30**:262–71.
30. Stuart JJ. Novel figurative language and patient experiencing in psychodynamic therapy. *Psychother Res* 1997;**7**:219–37.
31. Ulak BJ, Cummings AL. Using clients' artistic expressions as metaphor in counselling: a pilot study. *Can J Counsel* 1997;**31**:305–16.

F1
Family Engagement Questionnaire (FEQ)

General details	
Authors	Kroll L, Green J
Language	English
Country of publication/development	UK
Publication date	1997
Publisher	NA
Purpose and overview	
An instrument designed to enable clinicians to evaluate the therapeutic engagement of children and their families	
Theoretical orientation	Child/adolescent psychiatry
Population details	Child/adolescents
Perspective	Clinician/therapists
Measure used by	Child/adolescent clinicians
Other versions	No details
Notes	The FEQ was developed using a sample comprising three inpatient services: a forensic adolescent unit (Unit 1, $n = 7$), a regional adolescent service (Unit 2, $n = 13$); and a subregional child and adolescent unit (Unit 3, $n = 10$). 16 males and 14 females were in the sample, mean age 13.8 years
Areas of therapist–patient interaction addressed: Map	
Therapy context: boundaries; power/coercion	
Roles: friend/companion; confidant; protector	
Individual differences: level of functioning; problem complexity	
Therapist engagement: warmth; support/tolerance	
Patient engagement: motivation; attraction; commitment; intentions	
Framework: collaborative/participative/involving; structuring; directive; focused	
Threats to the relationship: intrusive; defensive; critical; hostility/anger; fear; resistance; confrontations	
Information derived from description of scales	
Dimensions	
Child–staff	Six items. The child's personal and therapeutic engagement with ward staff (intended to relate to the 'personal' aspect of treatment alliance)
Engagement with activities subscale	Four items. The engagement of the child with therapeutic activities on the ward (intended to relate to the child's 'task-related' or 'working' alliance)
Peer engagement	Three items. The child's personal engagement with other children in the milieu
Parental engagement	Three items. A measure of the parental engagement at both a 'personal' and a 'task-related' level with ward staff
Reliability	
For the measurement of internal consistencies, a split-half reliability assessment was made by dividing the total data set into two parts, A and B. The engagement with activities subscale showed adequate internal consistency as measured by Cronbach's alpha. The internal consistencies of child–staff and parental engagement showed partial internal consistency. The peer engagement subscale did not show internal coherence	
Inter-rater reliability was assessed by ICCs on the mean of data sets A and B. Only the parental engagement scale displayed adequate reliability; the other scales displayed partial reliability	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Split half	NA
Internal consistency	Child-staff: A 0.68, B 0.78 Engagement with activities: A 0.80, B 0.78 Peer engagement: A 0.33, B 0.13 Parental engagement: A 0.61, B 0.66
Inter-rater	Child–staff: 0.054 Engagement with activities: 0.57 Peer engagement: 0.59 Combined child scale (1+2+3): 0.63 Parental engagement: 0.73
Test–retest	No details
Validity	
Face and content validity issues have been addressed in the development of the questionnaire Only the peer engagement subscale displayed partial convergent validity with the independent clinical rating, as did the combined child subscales	
Face	The subscales discriminate between each other in a way that has face validity, i.e. the child subscales tend to intercorrelate together, but not to correlate with the parental subscale
Content	The FEQ has been developed out of clinical experience within inpatient child and adolescent psychiatry in line with recent theoretical work within the conceptualisation of therapeutic alliance in adult psychiatry
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	Correlations between the scales of the FEQ and the independent clinical rating scale were as follows: Child–staff: 0.07 Engagement with activities: 0.26 Peer engagement: 0.33 Combined child scale (1+2+3): 0.32 Parental engagement: 0.19
Discriminant	No details
Factor structure	The FEQ has four factors of child–staff, engagement with activities, peer engagement and parental engagement. Significant correlations are seen between the three children's subscales (range = 0.53 to 0.61), and there is a trend towards negative correlations between these children's subscales and the parental subscale, although these correlations do not reach significance
Responsiveness	
Discriminative (between individuals)	An ANOVA was used to compare subscale means of the FEQ and clinician rating against the variables of age, gender, and inpatient unit. No significant differences were found with gender or age. However, there was a variation according to the unit. Parents on Unit 1 engaged less well than did parents on Unit 2, with a trend towards significant difference in Unit 3. There were no differences shown in child alliance across the different units. The clinician instrument showed the same pattern of difference in parental alliance across the units
Evaluative (within individual across time)	No details

continued

Acceptability	
Number of items	16
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1997, Sage Publications
Web or scanning options	No details
Training details	No details
Administration/process details	The primary rating was completed by the key nurse of each patient, after consultation with others in the nursing team. A second rating was made by the co-working nurse attached to each patient
Support from measure developers	Copies of the questionnaire and details of coding are available from the first author
FAQ facility	No details
Precision	
Scale type	Likert visual analogue four-point scale, with higher scores indicating greater engagement
Normative data	See ref. 1 for FEQ scores across the three units
Notes	
One particular methodological issue needs to be taken into account when interpreting the psychometric properties of the FEQ. Because of the number of units involved in the study and their clinical organisation, a large number of staff was involved in the rating, totalling 30 nurses and eight clinicians. This was inevitable and indeed desirable since it meant that the instrument was tested in realistic clinical conditions, but such numbers are very likely to reduce inter-rater reliabilities and the correlations between clinician rating and the FEQ, making the instrument look less reliable than it should be if used in a single setting	
Résumé	
Strengths	The measure has been developed out of clinical experience within inpatient child psychiatry
Weaknesses	The reliability and validity of the FEQ is inconsistent. Not all of the scales meet criteria for adequacy
Areas for further research	Further testing of psychometric properties. Research on use and application in service settings
	Use of the measure in process research
Primary reference	
1. Kroll L, Green J. The therapeutic alliance in child inpatient treatment: development and initial validation of a family engagement questionnaire. *Clin Child Psychol Psychiatry* 1997;**2**:431–47.	
Secondary references	
None	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

F2
Family Therapeutic Alliance Scale (FTAS)

General details	
Author	Martin GR
Language	English
Country of publication/development	Australia
Publication date	1993
Publisher	NA
Purpose and overview	
A means of focusing on the emotional experiences of therapist and family	
Theoretical orientation	Pan-theoretical
Population details	See Notes
Perspective	Independent therapist
Measure used by	Family therapists
Other versions	36-item FTAS 24-item FTAS
Notes	*Practitioners:* Experienced family therapists[1] *Clients:* Families[1]
Areas of therapist–patient interaction addressed: Map	
Non-verbal communication: silence	
Therapist engagement: genuineness; respect; empathy/sensitivity	
Patient engagement: attraction	
Framework: convergent; collaborative/participative/involving	
Therapeutic techniques: responsiveness/receptiveness/attunement	
Threats to the relationship: critical; withdrawal	
Achieving a working relationship: working alliance	
Inferred from fully listed items: see Appendix 2 in ref. 1	
Dimensions	
Factors Alliance	Contains items on mutual liking and respect, good relationship, closeness–distance, the therapist being caring and well-joined
Joining	Items about joining with the family
Reliability	
In refining the scale, a criterion of partial adequacy was used in the measure of internal consistency	
Adequate inter-rater reliability and test–retest reliability of the FTAS were demonstrated	
Split-half	No details
Internal consistency	Communality is a measure of statistical association between items and suggests that items may be related to a common theme. Items not gaining a communality (alpha) of at least 0.55 were discarded[1]

continued

Inter-rater	Spearman rank correlation coefficients. Considering the reduced set of items, these correlations are high between the therapists (ranging between 0.77 and 0.90 at time 1 and 0.76 and 0.92 at time 2). This may mean that the scale is reliable in itself; conversely, it may represent the fact that team members had worked together for a time. Inter-rater reliability on the full original 36-item scale and on an interim 24-item scale was even higher[1]
Test–retest	Despite the small data set, the correlations for each therapist are high (range 0.78 to 0.95, all $p < 0.001$) and we can have some confidence that these are not chance events[1]
Validity	
Face, content and factor structure validity were addressed. Factor analysis of the original 24-item FTAS supports the construction of the present shorter 15-item FTAS	
Face	Questions were drawn from a review of all of the available published literature, particularly from work in the area of individual therapy. To these were added a range of questions based on empirical ideas considered by team members to be important to the notion of therapeutic alliance, or aspects of the process of the interview related to therapist or family functioning that might have some influence on therapeutic alliance. The remaining items appear to have face validity for therapeutic alliance, and take into account both therapist and family factors[1]
Content	Some idea of content and construct validity can be gained from consideration of the items contained in separate factors, their high item correlations and intercorrelations (factors 1 and 3 intercorrelate 0.60)[1]
Criterion (a) concurrent	No details
Criterion (b) predictive	While the family therapists who rated the families using the scale had no way of knowing the outcome of the families concerned, a prospective study is required to confirm what, at this point, may only be called an indication of support for the hypothesis that Family Therapeutic Alliance may be predictive of outcome in family therapy[1]
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	A principal components analysis with oblique primary pattern solution – varimax gave the best resolution of the 24 items into five factors accounting for 75.3% of the total variance. For the factor analysis it was assumed that each scale item had 100 responses (5 therapists × 10 videotapes × 2 occasions). Factors: 1 = alliance, 2 = lack of clarity, 3 = joining, 4 = family response, and 5 = shared view. Factor 2 had a moderately strong negative correlation with factor 1 (–0.58) and factor 3 (–0.54). Factors 3 and 2 have a moderately strong positive correlation (0.60). Factor 4 has a moderate positive correlation with factor 1 (0.42) and again with factor 3 (0.41). Factor 5 correlates best with factor 2, but then only weakly (0.26). Because of the strong positive correlation between factors 1 and 3 and the conceptual similarity between the ideas of joining and the bonding aspect of alliance, the authors decided to create a 'final' version of the FTAS which consists of only the 15 items from these two factors[1]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Responsiveness	
Discriminative (between individuals)	An ANOVA of the four global means gives an *F* test (3 df) = 19.61, $p = 0.0001$. Post hoc Scheffé analysis suggests that the alliance with families 1 and 2 was in each case significantly different from the alliance with families 3 and 4 at the 0.05 level, although the differences in alliance between families 1 and 2 or 3 and 4 did not reach significance. An ANOVA suggests that 13 of the 15 questions were able statistically to discriminate between alliances with *F* test values ranging from 5.77 to 25.6 and all with a very small probability of these being by chance ($p < 0.001$). The two families in sessions scoring highest completed therapy successfully. The family from the session scoring highest resolved their presenting problem in two sessions. The other family presented a symptom of recurrent and severe migraine and was able to gain symptom relief for the symptom bearer (among other changes in family dynamics) in 13 sessions. In contrast, the third and fourth families have made little change over a lengthy period and many years, respectively[1]
Evaluative (within individual across time)	No detail
Acceptability	
Number of items	15
Administration method	Rating scale
Time taken to complete	Limit of 15 minutes viewing time
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1993, Graham Martin and Stephen Alison
Web or scanning options	No details
Training details	No details
Administration/process details	Family therapists were given a brief overview of the development of the scale, and then viewed four 5-minute segments of videotaped family interviews, following each of which the scale was completed with no discussion[1]
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Judgements were made according to a seven-point time-sampling Likert scale from present 'all of the time' to present 'not at all'
Normative data	No details
Notes	
The first three drafts of the original scale were piloted during live family therapy sessions, with a family therapy team answering the questionnaire during the team break. Later, factor analysis reduced the number of items on the scale from 36 to 15. The measure was then tested with a group of family therapists	

continued

Résumé	
Strengths	Provided a useful focus for enhancing therapist–family relationships[1]
	Adequate internal consistency, inter-rater and test–retest reliability
Weaknesses	A limit of 15 minutes' viewing time was frustrating and the authors' particular way of choosing the time did not allow them to see beginnings (introductions and joining) and endings (termination of a session), both of which usually contain important clues about family members, therapist style and therapeutic alliance[1]
	Even if the scale can be shown to have reliability and validity, the issue still remains as to whose alliance is being measured. The authors have assumed throughout that the family is a unit, a system. A further assumption is that a global measure can apply to a family as a whole[1]
Areas for further research	While the process the team has been through suggests that family therapeutic alliance exists as a construct, which can be perceived and judged and can then be measured, the grounds on which the scale is developed, and therefore the validity of the construct and the scale, are shaky and in need of further confirmatory work and discussion[1]
	The use of the scale in a prospective study looking at family therapeutic alliance and outcome, a comparison of the therapists' view with the composite team's view, and adaptation of the scale so that a family may report on their own perceived alliance[1]

Primary reference

1. Martin GR, Allison S. Therapist alliance: a view constructed by a family therapy team. *Aust N Z J Fam Ther* 1993;**14**:205–14.

Secondary references

None

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

F3
Feminist Self-Disclosure Inventory (FSDI)

General details	
Authors	Simi NL, Mahalik JR
Language	English
Country of publication/development	USA
Publication date	1997
Publisher	NA
Purpose and overview	
The FSDI was developed to assess principles of therapist self-disclosure as described in the feminist therapy literature	
Theoretical orientation	Feminist therapy
Population details	No details
Perspective	Therapist
Measure used by	Feminist therapists
Other versions	No details
Notes	For the pilot study 150 feminist therapists and 150 non-feminist therapists (all female) were invited to participate by post. All respondents were asked to complete and return the FSDI. Ninety-one feminist therapists and 58 non-feminist therapists responded, giving a sample of 149 female therapists, mostly white. Mean age was 47.75 years and 76% held a doctoral degree. Years of experience ranged from less than 1 to 38
Areas of therapist–patient interaction addressed: Map	
Therapy context: confidentiality; boundaries; values; responsibilities	
Roles: advocate	
Therapist engagement: sensitivity; genuineness; respect; support/tolerance; openness; listening	
Framework: convergent; reciprocal; collaborative; congruent	
Therapeutic techniques: responsiveness/receptiveness/attunement; exploration	
Threats to the relationship: intrusive; critical; hostility/anger; resistance	
Outcomes: working alliance; emotional expression	
Information derived from example items	
Dimensions	
Therapist background (TB)	Five items, e.g. I inform my clients about my class background
Promotes liberatory feelings (PLF)	Four items, e.g. I believe self-disclosure can instill a sense of liberation in clients
Promotes egalitarianism (PE)	Four items, e.g. I use self-disclosure as an intervention with clients
Therapist availability (TA)	Three items, e.g. my clients know they may request for me to self-disclose in session and they do
Empowering clients (EC)	Three items, e.g. I inform my clients about my therapy orientation

continued

Reliability	
The FSDI total scale, and the TB, PLF and PE scales demonstrated adequate internal consistency as estimated by Cronbach's alpha. The EC scale demonstrated partial internal consistency[1]	
Test–retest reliability was estimated with Pearson's correlation at 2 weeks and 3 years. The total scale and the PE scale demonstrated adequate test–retest reliability at 2 weeks and none of the scales demonstrated adequate test–retest reliability at 3 years[1]	
Split-half	No details
Internal consistency	Total scale (0.88), TB (0.78), PLF (0.88), PE (0.80), TA (0.62), EC (0.54)
Inter-rater	NA
Test–retest	*2 weeks* Total scale (0.79), TB (0.54), PLF (0.69), PE (0.81), TA (0.40), EC (0.73) *3 years* Total scale (0.67), TB (0.74), PLF (0.52), PE (0.36), TA (0.58), EC (0.50)
Validity	
The FSDI demonstrated adequate content validity as measured by the consistency of the items with feminist principles/theory[1]	
The exploratory factor analysis demonstrated that the FSDI is composed of five factors accounting for a large proportion of the variance in the items. The intercorrelations of the factors indicated a moderate amount of shared variance[1]	
Construct validity of the FSDI was demonstrated by the support of the hypothesis that feminist therapists would endorse principles of feminist self-disclosure more than psychoanalytic/dynamic and other therapists. Other therapists refers to cognitive-behavioural, humanistic, family systems[1]	
Face	No details
Content	Four psychologists with expertise in feminist therapy were recruited to rate the 18 items on a ten-point Likert scale for their consistency with the principles and/or theoretical foundations of feminist therapy. Inter-rater reliability was 0.91 as measured by the ICC. Higher ratings represented more consistency with each item (1 = very inconsistent, and 10 = very consistent). The ratings ranged from 6.25 to 10
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	Feminist therapists scored significantly higher on the FSDI (M = 86.75) than other therapists (M = 75.30) and psychoanalytic/dynamic therapists (M = 66.87)
Convergent	No details
Discriminant	No details
Factor structure	A principal components exploratory factor analysis was conducted. Five factors emerged and were subjected to varimax rotation. The amount of variance accounted for by each factor was as follows: TB (0.34.6%), PLF (10.2%), PE (7.2%), TA (5.9%) and EC (5.6%). The correlations between the factors ranged from 0.34 to 0.56
Responsiveness	
Discriminative (between individuals)	The FSDI discriminated between therapists of a feminist orientation and therapists of other theoretical orientations (see construct validity section)
Evaluative (within individual across time)	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Acceptability	
Number of items	18
Administration method	Survey questionnaire
Time taken to complete	No details
Flesch reading age	NA
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1997, Cambridge University Press
Copyleft	Yes, in public domain
Web or scanning options	No details
Training details	No details
Administration/process details	The questionnaire was mailed to respondents. A cover letter described the procedure of the study and provided instructions for completing the questionnaire
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	A seven-point Likert scale was used, ranging from 1 = strongly agree to 7 = strongly disagree. To avoid acquiescent response bias, five of the 18 items were presented in a negative format so that agreement indicated opposition to that item. Higher scores indicate greater feminist self-disclosure. Minimum score = 18, maximum score = 126
Normative data	No details
Notes	
The generalisability of the findings is limited owing to use of an all-female, predominantly white sample	
The client's perspective is missing. This is particularly relevant as some items ask respondents to rate the impact of self-disclosure on clients (e.g. 'I believe my self-disclosure permits clients to validate their own feelings')	
The scale does not measure how often therapists use the intervention	
Scale items were not designed to measure inappropriate use of self-disclosure (e.g. over-disclosing)	
Résumé	
Strengths	Adequate internal consistency and test–retest reliability. Ability to respond between feminist and non-feminist therapists
Weaknesses	The FSDI is vulnerable to social desirability response bias. The client's perspective is missing. No items to measure inappropriate use of self-disclosure
Areas for further research	Use in training feminist counsellors
Primary reference	
1. Simi NL, Mahalik JR. Comparison of feminist versus psychoanalytic/dynamic and other therapists on self-disclosure. *Psychol Women Q* 1997;**21**:465–83.	
Secondary references	
None	

G1
Group Assessment of Interpersonal Traits (GAIT)

General details	
Author	Goodman G
Language	English
Country of publication/development	USA
Publication date	1972
Publisher	Jossey-Bass
Purpose and overview	
In the GAIT, participants are judged on the Rogerian (1957) constructs of empathy (accurate understanding), acceptance (warmth or unconditional positive regard) and openness (emotional honesty or genuineness)	
The GAIT assesses each applicant's (or group member's) solution to two problems: (1) how to go about disclosing an important part of one's self in far from ideal conditions; (2) how to enter into another person's frame of reference and understand his feelings with few questions and no judgements or interpretations or advice	
The traits that are rated are understanding, depressed, open, quiet, accepting–warm, rigid, relaxed and potential	
In the present version of Goodman's (1972) six-point scale, empathy is defined in terms of paying 'close attention', giving 'sensitive feedback', 'accurately understanding feelings as presented by the discloser', and taking care not to 'distract or interrupt the discloser's flow'[1]	
Theoretical orientation	Interpersonal therapy
Population details	See notes
Perspective	Peers (other group members), independent observer or self-report[3]
Measure used by	Designed to utilise peer raters with a minimum of technical rating experience, and so is appropriate for a variety of community mental health programmes. Useful as a selection device. Researchers
Other versions	Group Assessment of Interpersonal Traits – 1974
Notes	*Practitioners:* Applicants for a training programme. The applicants participated in the GAIT procedure as part of the selection process. Some were rejected, and the data pertain to those who were accepted and participated in GAIT procedures as part of their training. Participants are referred to interchangeably as applicants and students[3] College undergraduates, mature housewives seeking a second career as a paraprofessional counsellor[2]
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: empathy/sensitivity; warmth; listening	
Framework: flexible/rigid	
Therapeutic techniques: responsiveness/receptiveness/attunement; feedback	
Inferred from GAIT items as listed in ref. 3.	
Dimensions	
Understanding	I feel he understands what others really mean
Depressed	He seems sad, blue, discontented
Open	He appears honest, frank, emotionally open
Quiet	I see him as a mild, reserved quiet person
Accepting–warm	He seems warm, patient and accepting

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Rigid	He appears set in his ways
Relaxed	I see him as a relaxed, easy-going person
Therapeutic talent	A composite of 'understanding', 'open' and 'accepting–warm'
Reliability	
The split-half reliability of the GAIT has been found to vary from being inadequate to adequate, between studies and items	
The level of internal consistency varies as a function of the position of the rater, with the peer raters producing more adequate reliabilities. Intercorrelations of the items demonstrate inadequate to adequate reliability	
Findings for the GAIT demonstrate inadequate to partially adequate inter-rater reliability	
The mean test–retest reliability is adequate	
Split-half	Observer groups of four to five combined staff and students. Reliabilities were found to be from 0.44 to 0.79. The section also reports the split-half reliabilities for the therapeutic talent items from the Goodman study, followed by the findings of Chinsky and Rappaport (1971) as follows: understanding 0.64, 0.70; accepting–warm 0.63, 0.41; open 0.54, 0.56. Mean reliabilities for the three items from the two studies are 0.60 and 0.56[3]
Internal consistency	Peer GAIT variables were more highly intercorrelated than trained GAIT variables. The two understander role variables, empathy and acceptance, were significantly and positively related for both peer ($r = 0.75$) and trained GAIT raters ($r = 0.69$, $n = 130$, $p < 0.001$). Openness was significantly related to empathy and acceptance for peer raters ($r = 0.30$ for both, $p < 0.001$), but not for trained raters ($r = 0.01$ and 0.02, respectively). Within a GAIT dyad, the quality of the understander's behaviour seemed intertwined with the openness of the discloser[2]
	In general, most of the findings on GAIT suggest a coherent internal order among the GAIT items. A table is presented of the intercorrelations of the GAIT items as evidence for its internal consistency (values range from –0.61 to –0.06 and from 0.02 to 0.84). Correlation patterns are reported to make sense and make designs that fit intuitive expectations based on item definitions. None of the correlations jarred expectations, although some intercorrelations approached the limits set by their reliabilities so that the measures are not always sharply distinguished from each other[3]
Inter-rater	Inter-rater correlations: peer GAIT empathy 0.45, trained GAIT empathy 0.45. Ebel intraclass correlation of 0.68. Using Kendall rank order correlation, peer GAIT empathy and trained GAIT empathy correlated 0.50 ($p < 0.001$). Untrained raters tended to agree with the more highly trained raters within each definition and session[1]
	Average Spearman–Brown corrected reliabilities for the three trained raters = 0.66. Peer and trained GAIT ratings were, as in previous studies, positively correlated. Peer ratings consisting of all but self-ratings were significantly related to trained ratings for each variable (r ranged from 0.45 to 0.50). Self-ratings were significantly and positively related to both group peer and trained ratings for all three GAIT variables. Self-ratings of acceptance had the best agreement with others' ratings ($r = 0.30$ with group peer and 0.41 with trained GAIT acceptance, $n = 124$, $p < 0.001$). Trained GAIT raters typically gave fewer high scores than peer raters did. Absolute GAIT ratings generally declined as the point of view of the rater moved from the rated person himself to his dyadic partner, to the non-participating observers. Goodman reports a study of his own with 180 student participants where inter-judge reliabilities between three experienced staff raters were computed. The mean coefficient (Spearman–Brown) on all GAIT items was 0.51, and the mean coefficient on the three therapeutic talent items was 0.45. The section reports mean coefficient reliabilities between advanced clinical psychology graduate students for the three therapeutic talent ratings from mixed-gender GAIT groups of 0.52 (Chinsky and Rappaport, 1971) and 0.52 (D'Augelli *et al*., 1971)[2]
	continued

	Students ($n = 180$) participated in the GAIT procedure and were rated by each other and by staff members. Students did not rate themselves. Correlations of student and staff ratings on 179 applicants ranged from 0.23 to 0.52, which were modest but significant at the 0.01 level. Inter-judge reliability for the three staff raters was computed on all GAIT items and produced a mean coefficient of 0.51 (Spearman–Brown correction)[3]
Test–retest	41 male and female undergraduates took GAIT on two occasions, approximately 3 weeks apart. The study was confined to students' ratings; no external judges were involved. Coefficients ranged from 0.66 to 0.86, with a mean of 0.80 (Dooley, 1972)[3]
Validity	
The GAIT's validity has been demonstrated in both field and laboratory studies (Dooley, 1975)	
Trained GAIT empathy has partial predictive validity for the criterion of counselling readiness at 9-month follow-up. Predictive validity ranges from being inadequate to partial with regard to measure of improvement	
Significant differences were found between medium and high scores on the GAIT, which the authors[3] claim demonstrates construct validity	
Inadequate to partial convergent validity has been demonstrated for the GAIT	
Face	No details
Content	The best combination of trained GAIT empathy and self-reported experience in counselling had a multiple correlation of 0.63 with the criterion of counselling readiness. When the more laborious staff ratings were used instead of self-reported counselling experience in combination with trained GAIT empathy, a multiple correlation of 0.66 was obtained with the criterion. Only trained GAIT empathy correlated with the 9-month follow-up ratings of counsellor readiness (Kendall 0.40, Pearson 0.48, $p < 0.01$)[2]
	Associations were calculated for four variables of change in emotionally troubled boys and GAIT scores of their college student companions. In general, the correlations fell in a systematic pattern and lend support to GAIT as a predictor of therapeutic talent in a field situation. Of the 36 correlations, 32 fall in directions suggesting associations between GAIT scores and boy improvement. However, the correlations do not indicate that the predictors are powerful (range –0.02 to –0.31, 0.1 to 0.26). The section reports five studies which all offer support for the predictive validity of GAIT therapeutic talent items. One such study is Rappaport *et al*. (1971), who included the GAIT in several pretherapy procedures designed to predict therapeutic ability in 36 student volunteer group leaders for hospitalised schizophrenic patients. Observer-rated GAIT acceptance–warmth was significantly associated with the staff-rated Ellsworth Behavioral Adjustment Scale: improved mood (0.39, $p < 0.05$); cooperation (0.41, $p < 0.05$) and total adjustment (0.46, $p < 0.01$). GAIT understanding correlated with improved mood (0.48). No student-rated GAIT variables correlated significantly with outcome measures[3]
	Ratings utilising more similar construct definitions (Carkhuff empathy and Carkhuff gross) agreed more than ratings based on dissimilar constructs (i.e. GAIT and Carkhuff)[1]
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	The accepted candidates' GAIT data were grouped into medium and high scores for each GAIT item. Medium and high scorers were compared on relevant external variables to see whether differences suggested any construct validity. High scorers on GAIT accepting–warm differed significantly from low scorers in several respects: they chose significantly ($p < 0.01$) more person-orientated vocational goals; they described themselves as less dominant, exhibiting and self-confident ($p < 0.01$); they scored higher on the Adjective Check List (ACL; Gough and Heilbrun, 1965) deferent ($p < 0.05$) and lower on the ACL defensive ($p < 0.10$); 5 months later they were described as letting

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	others be themselves, made fewer attempts to influence others and were rated as less assertive and determined (0.05). The section reports extensive findings following the same procedure which offer similar support (correlations with relevant variables) for the GAIT constructs of quiet, understanding others' feelings, rigid and potential for the counsellor role. Quiet and understanding others' feelings were correlated with existing published measures. Few relevant differences were found between high and moderately high scorers on the blue and relaxed scales[3]
Convergent	Peer GAIT empathy did not significantly correlate with peer or trained Carkhufff empathy, or with trained Carkhuff gross rating of facilitative functioning in session 1. It did correlate 0.25 ($p < 0.05$) with trained Carkhuff Gross as measured in session 2. Similarly, trained GAIT empathy only significantly correlated with trained Carkhuff gross as measured at session 2 (0.30, $p < 0.01$). The relationship between GAIT empathy and Carkhuff accurate empathy was negative and not significant (–0.10). It appears that at least some of the difference between GAIT and Carkhuff empathy ratings can be closed by employing (1) a construct which falls between the Carkhuff and GAIT empathies on the continuum ranging from specific to global, and (2) the same situational sample[1]
	The first impression variable predicted subsequent peer GAIT ratings of empathy ($r = 0.55$) and acceptance ($r = 0.43$, $N = 130$, $p < 0.001$), but not openness ($r = 0.13$, $p < 0.10$). First impression scores also correlated with trained GAIT empathy ($r = 0.15$, $p < 0.05$) and acceptance ($r = 0.21$, $p < 0.01$). The sum of peer empathy, acceptance and openness correlated 0.27 with staff (made without knowledge of the GAIT results) and 0.41 with selection, but only 0.33 with selection when staff was held constant by partial correlation[2]
	Participants' self-descriptions of quiet or reserved versus outgoing were strongly related to the GAIT quiet score, in the appropriate direction. There was also a positive correlation between GAIT quiet scores and low participation in extracurricular activities ($p < 0.05$). The 25 scales of the ACL were correlated with the GAIT items. The ACL exhibit scale (adjectives such as outspoken) showed the strongest correlation with GAIT quiet (–0.28, $p < 0.01$); GAIT blue correlated with ACL unfavourably (0.20, $p < 0.01$); GAIT therapeutic talent negatively correlated with ACL defensive (–0.20, $p < 0.01$)[3]
Discriminant	No details
Factor structure	NA
Responsiveness	
Discriminative (between individuals)	Other things being equal, ratings based on the same sample will be more similar than ratings based on different samples (therefore, there was only limited agreement between sessions 1 and 2)[1]
	The reflection subjects gave fewer high empathy ratings (49%) than did the control Ss (61%, $t = 2.86$, $n = 111$, $p < 0.005$). However, this apparent training effect did not appear for the other GAIT variables[2]
	See construct validity[3]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	Seven
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

Feasibility	
Copyright	1972, Jossey-Bass
Web or scanning options	No details
Training details	Important to standardise the intimacy level of GAIT disclosure and to train raters to weigh the relative contributions of both participants in a GAIT dyad. Some of the participants in this study participated in a 45-minute pre-GAIT training experiment[2] The 'trained raters' in ref. 2 were upper-level undergraduates who each spent approximately 40 hours in preparation using the author's training manual (Dooley, 1973) and previously rated audio-recordings of pilot GAIT sessions
Administration/process details	The GAIT procedure consists of a series of 5-minute discloser–understander dyads followed by evaluations by the participants (peer ratings) or by observers (in person or subsequently from audio- or video-recordings). Each participant takes each role once. Each participant is asked to write two disclosures about current interpersonal concerns. In the discloser role, the participant was directed to read the more difficult (interpersonally risky) disclosure if possible, but could read the less difficult one if he felt too uncomfortable, while the understander is asked to show understanding to the discloser. Participants or applicants are rated by the other group members and by attending staff members Audio-recordings of the GAIT sessions can be subsequently rated by trained raters on the same scales Peer and trained GAIT ratings were computed as the percentage of raters giving Ps a high rating (3 or 4 on the four-point scales)[2]
Support from measure developers	The GAIT procedure, scoring method and score patterns are detailed for those who wish to experiment with the procedure in ref. 3
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert. Each statement, e.g. 'I feel he understands what others really mean', is rated on a scale of six, ranging from 'very much like him' to 'very much not like him'. The instrument also contains a space for rank-ordering the applicants on judged potential as successful counsellors. The item 'Therapeutic talent' is a composite of the accepting, understanding and open items[3]
Normative data	No details
Notes	
Relative to Carkhuff empathy, the GAIT empathy scale encompasses more behavioural elements and is applied with more rater discretion to whole 5-minute segments of interaction. While somewhat less reliable, the GAIT empathy rating procedure is typically used to assess larger units of interaction which may be more representative of average expectable behaviour	
Résumé	
Strengths	Mean test–retest reliability is adequate. GAIT demonstrates responsiveness Trained GAIT empathy has partial predictive validity for the criterion of counselling readiness at 9-month follow-up. Significant differences were found between medium and high scores on the GAIT, which the authors[3] claim demonstrates construct validity
Weaknesses	Inter-rater, split-half and internal consistency reliability values vary widely from being inadequate to adequate Predictive validity ranges from being inadequate to partially adequate with regard to measure of improvement

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	Inadequate to partially adequate convergent validity has been demonstrated for the GAIT
	Some of the criteria and external variables that are used to assess the GAIT's validity may be considered as not very rigorous
	Lengthy training procedure
Areas for further research	Further research into psychometric properties to reduce disparity in findings

Primary references

1. Dooley D, Lange AJ, Whiteley JM. Sources of discrepancy in Carkhuff and Gait measurements of empathy. *Psychother Theory Res Pract* 1979;**16**:337–44.
2. Dooley D. Selecting nonprofessional counselor trainees with the group assessment of interpersonal traits (GAIT). *Am J Community Psychol* 1975;**3**:371–83.
3. Goodman G. *Companionship therapy: studies in structured intimacy*. San Francisco, CA: Jossey Bass; 1972.

Secondary references

4. Chinsky JM, Rappaport J. Evaluation of a technique for the behavioral assessment of nonprofessional mental health workers. *J Clin Psychol* 1971;**27**:400–2.
5. Dicken C, Bryson R, Kass N. Companionship therapy: a replication in experimental community psychology. *J Consult Clin Psychol* 1977;**45**:637–46.
6. Dooley D. Preliminary validation of a nonverbal description form for help-intended interactions. *Psychother Theory Res Pract* 1978;**15**:140–4.
7. Kramer JA, Rappaport J, Seidman E. Contribution of personal characteristics and interview training to the effectiveness of college student mental health workers. *J Counsel Psychol* 1979;**26**:344–51.
8. Levant RF, Geer MF. A systematic skills approach to the selection and training of foster parents as mental health paraprofessionals: I. Project overview and selection component. *J Community Psychol* 1981;**9**:224–30.
9. Lindquist CU, Rappaport J. Selection of college student therapeutic agents: further analysis of the 'group assessment of interpersonal traits' technique. *J Consult Clin Psychol* 1973;**41**(2).
10. Teevan KG, Gabel H. Evaluation of modeling-role-playing and lecture-discussion training techniques for college student mental health professionals. *J Counsel Psychol* 1978;**25**:169–71.

H1
Helper Behaviour Rating System – Modified Version

General details	
Author	Shapiro D (author of original version Elliot R)
Language	English
Country of publication/development	UK (original version USA)
Publication date	1984 (original version 1979)
Publisher	NA
Purpose and overview	
A scale for response mode analysis, which analyses the interpersonal function served by single units of therapist or client speech	
NB. This scale is a modification of the Helper Behaviour Rating System (Elliott, 1979);[3] see Other versions, below	
Theoretical orientation	Person-centred/pan-theoretical
Population details	Clinical students: 12 polytechnic and university students, two male, two female, seeking help with personal and relationship problems[1]
Perspective	Independent rater.[1] Raters were the second author (clinical psychologist) and two postgraduate students. Twelve 50-minute counselling sessions were rated
Measure used by	Counsellor: three student counsellors (client-centred) with at least 7 years' experience each saw two male and two female clients
Other versions	This scale is a modification of the Helper Behaviour Rating System (Elliott, 1979),[3] with the addition of a category 'exploration', which is intermediate between Elliott's 'interpretation' and 'reflection'. The 'exploration' category or mode taps the effort by the helper to construct a frame of reference shared with the client
Notes	
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: hope/encouragement; openness; listening, praise/affirmation; empathy/sensitivity (developing the relationship)	
Therapeutic techniques: exploration; reflection in action; feedback (maintaining the relationship)	
Information derived from description of dimensions	
Dimensions	
Categories are	Categories rather than dimensions: 11 categories plus an 'other' category for units not codable in any category
Exploration	Contains two subtypes: (i) inside: comprises responses describing feelings or thoughts going on in the client but which the client has not yet verbalised. Corresponds to Elliott's (1979) interpretation (inside) category; (ii) reformulation: comprises Elliott's (1979) reflection (implication) which includes responses verbalising content implied by the client, plus a subset of responses from Elliott's interpretation (classifying) category, which label an experience without offering diagnosis or judgement
Closed question	Gathering restricted information
Open question	Gathering unrestricted information
General advisement	Advising the client to some action outside the therapy session itself
Process advisement	Advising the client to some action within the session
Reflection	Re-presenting the client's message

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Interpretation	Giving new information about the client
Reassurance	Responding positively to the client
Disagreement	Responding negatively to the client
Self-disclosure	Significantly revealing the therapist
Information	Giving new information not about the client
Reliability	
Reliability is adequate, although the results obtained for some categories are sufficiently modest to require caution in their use	
Split-half	NA
Internal consistency	NA
Inter-rater	The overall kappa between the three raters was 0.76 and the pairwise kappas between the second author and the other two coders were 0.78 and 0.79, and the kappa between the other two coders was 0.72.[1] Overall, 74.3% of the 4583 units were coded unanimously by all three coders, and a further 22.9% were coded similarly by two of the three coders. Kappas for each category are presented in ref. 1 and range from 0.97 (reassurance) to 0.53 (exploration) All coders coded the four interviews involving one counsellor and then met to resolve discrepancies before moving onto the next counsellor
Test–retest	No details
Validity	
There is no relevant validity information on the categories that comprise this measure, other than that it is a development of an existing measure by Elliott (1979)[3]	
Face	No details
Content	Addressed (no further details available)
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	NA; 12 categories
Administration method	Independent raters rate therapy transcripts (pencil and paper)
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

Feasibility	
Copyright	Shapiro *et al*., 1980, University of Sheffield[2] Original version: *Journal of Counseling Psychology* (Elliot, 1979)[3]
Web or scanning options	No details
Training details	The main rater had 2 years' research experience with response mode analysis, and trained other coders in the study.[1] This training lasted approx. 20 hours, including presentation of the modified coding manual,[2] coding of practice examples and five practice transcripts
Administration/process details	Counselling sessions were recorded and transcribed.[1] Raters use this information and code therapist utterances and behaviours using the 12 categories of the measure
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Nominal, binary
Normative data	No details
Notes	
The original version (Elliott, 1979)[3] has a hierarchical arrangement of 47 subtypes into ten major categories; the subtypes are used to aid the coder in assigning the utterances to a major category	
Résumé	
Strengths	Relevant for UK clinical research. Reliability overall is acceptable. The modified version achieves greater precision than the original
Weaknesses	The reliability results obtained for some categories are sufficiently modest to require caution in their use
Areas for further research	Modest reliabilities of the 'interpretation' and 'exploration' codes suggest a need for further refinement of the coding manual, e.g. to provide more examples to delineate the boundaries of these codes
Primary references	
1. Shapiro DA, Barkham M, Irving DL. The reliability of a modified Helper Behaviour Rating System. *Br J Med Psychol* 1984;**57**:45–8. 2. Shapiro DA, Barkham M, Irving DL. A modified Helper Behaviour Rating System. SAPU Memo No. 415. University of Sheffield; 1980.	
Secondary reference	
3. Elliott R. How clients perceive helper behaviours. *J Counsel Psychol* 1979;**26**:285–94.	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

H2
Helpful Responses Questionnaire (HRQ)

General details	
Author	Miller WR
Language	English
Country of publication/development	USA
Publication date	1991
Publisher	No details
Purpose and overview	
Designed as an open-response questionnaire for individual or group administration, analogous to the Truax scale, and conceptually linked to Gordon's (1970) description of active listening. Developed as part of a project to develop, implement and evaluate training materials for crisis intervention counsellors in rural community settings[1]	
Theoretical orientation	No details
Population details	See below
Perspective	Therapist self-report
Measure used by	Practitioners
Other versions	No details
Notes	*Practitioners:* Paraprofessional *n* = 120. Average age 37.3, mean of 14.4 years of education, 109 white non-Hispanic Of the 190 who began training, 120 completed the HRQ both before and after the workshop[1] *Rater:* Research assistant
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: empathy/sensitivity	
Therapeutic techniques: responsiveness/receptiveness/attunement; reflection in action	
Inferred from description of scoring guidelines in ref. 1	
Dimensions	
No details	
Reliability	
The internal consistency of the HRQ demonstrates partial to adequate reliability	
Adequate inter-rater reliability of the HRQ has been demonstrated	
The test–retest reliability of the HRQ is inadequate	
Split-half	No details
Internal consistency	The mean inter-item correlation was 0.67 at pre-training and 0.57 at post-training. Cronbach's alpha statistic was 0.92 at pre- and 0.89 at post-training. Mean correlations between item scores and total score (corrected by removing the item being correlated) were found to be 0.87 and 0.79, respectively[1]
Inter-rater	Pearson product moment. Reliability coefficients for individual items ranged from 0.71 to 0.91 (all $p < 0.001$). The reliability of the principal raters when checked against a trainer on 120 randomly chosen responses proved to be 0.85 and 0.83 ($p < 0.001$). The inter-rater reliability for the total HRQ scores (sum of the six item scores for each respondent) was 0.932 ($p < 0.001$)[1]
	continued

Test–retest	Evaluated test–retest reliability by correlating trainees' two scores, recognising that a period of training intervened between first and second testing. Combined with the significant training effect, this modest coefficient (0.45) suggests differential skill acquisition, such that some individuals showed improvement in empathy, whereas others did not[1]
Validity	
The HRQ was found to correlate significantly with a self-esteem inventory, but no figures were provided, and so conclusions as to the convergent validity of the HRQ are limited. No other areas of validity have been addressed	
Face	Scale definitions integrate Truax's depth rating system with concepts from Gordon (1970)
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	Self-esteem (as measured by self-esteem inventory, Coopersmith, 1975) was found to be related positively to empathy scores, $r_{190} = 0.19$, $p < 0.01$[1]
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	A highly significant main effect of training was found ($F_{1,118} = 101.2$, $p < 0.001$), an indication that training produced substantial improvement as reflected on the HRQ[1]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	6
Administration method	Questionnaire
Time taken to complete	Average administration time is 15–20 minutes
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1991, Clinical Psychology Publishing
Web or scanning options	No details
Training details	No details
Administration/process details	Consists of six paragraphs that simulate communications from individuals with specific concerns. After each paragraph, a space is provided for the respondent to write a helping response. HRQ was administered at the beginning of the first workshop day before any training and at the end of the second day. Blind rating system used[1]
	Individual or group administration
Support from measure developers	Items and instructions included in ref. 1
FAQ facility	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Precision	
Scale type	The HRQ is scored by rating each response on a five-point ordinal scale of depth of reflection. A rating of 1 is assigned when the response contains no reflection, but does include at least one element scoreable as a 'roadblock' response as defined in Gordon's (1970) 'typical twelve' responses. A 5 is scored when the response qualifies at level 4 and also includes either a reflection of feeling that fits the original statement or an appropriate metaphor or simile
Normative data	Responses at level 1 or 2 can also be scored further to indicate which of Gordon's 12 roadblocks they contain (Miller and Jackson, 1985). This can be useful as training feedback[1]
	Normative data are reported based on a sample of 190 paraprofessional trainees[1]
Notes	
Workshops designed to train paraprofessional crisis intervention counsellors were offered in 14 rural New Mexico communities. Content of the workshops included 6–8 hours of training in active listening skills and 6–8 hours of instruction and practice in other crisis intervention skills, distributed over 2 days. Two junior authors served as trainers[1]	
Résumé	
Strengths	The internal consistency of the HRQ demonstrates partial to adequate reliability
	Adequate inter-rater reliability of the HRQ has been demonstrated
	The HRQ is an alternative when individual observation is not feasible, as when groups are being assessed and trained[1]
	Potential for training feedback (see 'scale type')
Weaknesses	The test–retest reliability of the HRQ is inadequate. Very limited amount of psychometric data available on the HRQ
Areas for further research	Other unmeasured variables probably account for variance in empathic skills, as defined by the HRQ, and further study is needed to clarify correlates of trainees' initial level and acquisition of empathy[1]
	Future studies could explore the convergence of questionnaire and observational data and the stability of training-related changes in therapeutic empathy[1]
Primary reference	
1. Miller WR, Hedrick KE, Orlofsky DR. The Helpful Responses Questionnaire: a procedure for measuring therapeutic empathy. *J Clin Psychol* 1991;**47**:444–8.	
Secondary references	
None	

H3
Helping Alliance Counting Signs Method (HAcs)

General details	
Author	Luborsky L
Language	English
Country of publication/development	USA
Publication date	1983
Publisher	NA
Purpose and overview	
To extend the applicability of clinicians' ratings of the helping alliance through identifying specific clues of existence in transcripts from psychotherapy sessions[4]	
Theoretical orientation	Been applied to individual, support-expressive psychoanalytically orientated psychotherapy[1,3]
Population details	See below
Perspective	Independent judges
Measure used by	Practitioners, researchers
Other versions	Helping Alliance Rating Scale
Notes	*Clients:* Normative sample: the ten most and the ten least improved among the 73 patients in the Penn Psychotherapy Project. The demographic characteristics of the ten most and ten least improved patients were similar; all were non-psychotic patients, 13 female, mean age 26[1] All patients were non-psychotic. The demographic details characteristics of the ten most (MI) vs the ten least (LI) improved patients were similar to each other; the mean age was 26, and 15 were female[3] Depressed patients (Luborsky *et al*., 1999)[2] *Practitioners:* Psychotherapist[3] Paraprofessional[1,3] Therapist[1] Psychiatrist[3] *Rater:* Clinically experienced[4]
Areas of therapist–patient interaction addressed: Map	
Therapy context: values	
Individual differences: level of functioning	
Therapist engagement: respect; support/tolerance	
Patient engagement: motivation; expectation/preferences; attraction	
Framework: collaborative/participative/involving; convergent	
Outcomes: general satisfaction; achieving a working relationship; changing view of self with others	
Inferred from the HAcs manual and the information provided below	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Dimensions	
	The manual consists of seven subtypes of two broad types of patients' helping alliances together with examples
Perceived helpfulness of the therapist (HA type 1)	(1) The patient believes that the therapy is helping (2) The patient feels changed since the beginning of the treatment, or he is considered to be better (3) The patient feels a rapport with the therapist, and feels understood and accepted (4) The patient feels optimism and confidence that the therapist and treatment can help
Patient's collaboration with the therapist (HA type 2)	(5) The patient experiences the treatment as working together with the therapist in a joint effort, as part of the same team (6) The patient shares with the therapist similar conceptions of the aetiology of the problems (7) The patient demonstrates qualities that are similar to those of the therapist, especially in having the tools for understanding
Reliability	
The inter-rater reliability of the HAcs ranges from being partial to adequate	
Split-half	NA
Internal consistency	NA
Inter-rater	The reliability estimate for the HAcs method is more complex and the results more mixed than for the Help Alliance Rating Method (HAr) method. While we would conclude that the inference process is simpler in the HAcs method than in the HA method, this conclusion only applies to the scoring of the same statements[1] Kappa of 0.80 ($p < 0.05$) (Luborsky *et al*., 1999, sample of 19 depressed patients using an enlarged manual to increase the range of the measure)[2] Expressed correlationally, the positive items were agreed on especially well (for early sessions 0.69, for late sessions 0.82, $p < 0.001$). Only 74 signs among all those scored by both judges were assigned to exactly the same statements. However, agreement on all components simultaneously is well beyond the call of duty – very high agreement was obtained for each component alone: for value (100%), for type (97%), for subtype (82%) and for intensity (61%). Two types of unreliability in the HAcs approach must be distinguished (see Notes). Many of the judges' errors were of the first type and did not involve 'true unreliability'[3,4]
Test–retest	No details
Validity	
Adequate content, concurrent, predictive and construct validity have been demonstrated for the HAcs	
The convergent validity of the HAcs is adequate, particularly regarding the positive signs, the difference between positive and negative, and in the later rather than earlier sessions	
Face	No details
Content	The scale types and subtypes were suggested by inspection of preliminary sessions, as well as by concepts of the helping alliance in clinical writings[2,4]
Criterion (a) concurrent	Support for the concurrent validity of the scales comes from examining a number of correlations derived both from during-treatment and pretreatment measures. Only example: early HAcs positive signs were not significantly correlated with early Therapist Facilitating Behaviors Counting Signs Method (TFBcs) positive signs. However, late HAcs positive signs were correlated 0.80 ($p < 0.001$) with late TFBcs positive signs[1]

continued

Criterion (b) predictive	Expressed correlationally, the HAcs attained moderately successful predictions, with correlations of around 0.5 on four major outcome measures: rated benefits; residual gain, a summary measure of the therapist's view of the patient's degree of success, satisfaction and improvement (SSI) during treatment; and change in target complaints[1] The positive signs were more frequent and more predictive by far than the negative ones (no figures provided). The positive signs therefore were more reliable harbingers of eventual beneficial outcomes of the treatments (Luborsky *et al*., 1988)[2] Significant predictive power (examined by two-factor mixed model ANOVA and by correlations) was found for HAcs. The highest correlations tended to appear for the early positive HAcs; it was correlated with a rated benefits measure 0.57 ($p < 0.01$) and with a residual gain measure 0.58 ($p < 0.01$). The combination of three simple outcome rating scales by the therapist – success, satisfaction and improvement (SSI) – was significantly predicted by early positive HAcs (0.59, $p < 0.01$). Early positive HAcs correlated 0.59 ($p < 0.05$) with change in the first target complaint (specific symptoms for which the patient came for treatment)[3,4]
Construct	Late HAcs positive was correlated with late TFBcs positive, 0.80 ($p < 0.001$). This may suggest that the therapists' attempts to facilitate HA behaviour eventually (i.e. late) became successful[3]
Convergent	HAcs and HAr were significantly correlated for both early and late session ratings (range from 0.57, $p < 0.01$, to 0.86, $p < 0.001$, for positive signs and difference between positive and negative)[1,2] They agreed more highly for the late sessions (0.83, $p < 0.001$) than for the early sessions (0.57, $p < 0.01$), perhaps because in the late sessions the outcome of the treatment might have been more evident in what the patient and the therapist said in the session. Luborsky (1999): in the depressed patient sample using an enlarged manual, the correlation of the HAcs with the HAr was only low to moderate, 0.51 (significant at the 0.05 level, two-tailed)[2] The greater agreement for HAcs positive signs (range 0.57, $p < 0.01$, to 0.86, $p < 0.001$) than for negative signs (range –0.14 to –0.21, ns) may be due in part to the fact that there were fewer negative signs[3,4] Authors looked at the correlates on HAcs with pretreatment measures. The similarities of patient and therapist score was significantly correlated with positive HAcs (0.60, $p < 0.01$), and the difference between positive and negative signs (0.62, $p < 0.01$). The HAcs difference score (0.47, $p < 0.05$) and HAcs negative signs (–0.61, $p < 0.01$) significantly correlated with other ratings by members of staff who had known the therapists' work over the years. Therapists' Embedded Figures Test (Witkin, 1949) correlated with early positive HAcs 0.62 ($p < 0.001$), and with HAcs difference score 0.51 ($p < 0.05$). The Health Sickness Rating Scale (HSRS) correlated with positive HAcs 0.44 ($p < 0.05$)[4]
Discriminant	No details
Factor structure	Although the dimensionality of the HAcs and the HAr methods has not yet been directly evaluated by means of factor analysis, other validation efforts have been undertaken for each of the instruments[1]
Responsiveness	
Discriminative (between individuals)	For the HAcs method, there were significant *F* ratios for the outcomes for positive (16.3, $p < 0.001$) and difference scores (7.1, $p < 0.05$), meaning that the more improved group had more HAcs signs than the less improved group[1]
Evaluative (within individual across time)	Helping alliance scores of the two early sessions are moderately consistent with scores of the two late sessions for the HAcs method (positive items) ($r = 0.58$, $p < 0.01$)[3,4] For the more improved group the correlations were: by the HAcs method, positive 0.54; negative 0.47; difference between positive and negative 0.47 ($p < 0.05$)[4]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Acceptability	
Number of items	NA
Administration method	Judge rating scale
Time taken to complete	Each transcript consisted of the first 20 minutes (or at least ten typewritten pages) from four psychotherapy sessions: sessions 3 and 5 and two late sessions[4] Since the HAr method is less time-consuming than the HAcs method, it has a practical advantage[3]
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1983, *Journal of Nervous and Mental Disease*
Web or scanning options	No details
Training details	It is essential that raters be clinically experienced to use the HA methods
Administration/process details	Each of the seven subscales may appear in positive or negative form. First, the judge is required to locate all relevant patient statements, that is, 'signs', in the transcript that fit each helping alliance subtype, to classify them as positive or negative, and then to rate their intensity on a five-point scale (from 1 = very low, to 5 = very high). Each patient's score is the sum of the number of signs in each session, weighted by the intensity ratings[4]
Support from measure developers	Manual in Appendix A of Luborsky (1976) and Luborsky (1983)
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert (from 1 = very low, to 5 = very high)
Normative data	No details

Notes

The HAcs has been used much less than the HAr in later research. It may be that the expected labour of using the HAcs discourages some researchers from trying it. However, although the HAr seems simpler, the evidence so far is that the HAcs is not difficult to use and the psychometric properties of the HAcs appear to be somewhat better than those of the HAr[2]

Although time-consuming to use, the HAcs method does have the beneficial quality of a precise location for the content scored[1]

A combination of the two main types of procedures should be used for assessment of the alliance: a self-rating questionnaire method (e.g. the HAQ-II, the revised and expanded alliance measure) and a clinical observer rating method (e.g. the HAcs or HAr)

Two types of unreliability in the HAcs approach must be distinguished: (1) when one judge does not score a unit of the transcript which another judge had located and specified as fitting a subtype of the counting signs manual; this 'locational unreliability' can sometimes be ascribed to lack of attention by one judge to the particular statement and may not be as serious as 'true unreliability'; (2) when two judges assign different counting signs subtypes to the same portion of the transcript; this might be referred to as 'true unreliability'. Many of the judges' errors were of the first type and did not involve 'true unreliability'[3,4]

continued

Résumé	
Strengths	The inter-rater reliability of the HAcs ranges from being partially to adequately demonstrated
	Adequate content, concurrent, predictive and construct validity have been demonstrated for the HAcs. The convergent validity of the HAcs is adequate, particularly regarding the positive signs, the difference between positive and negative, and in the later rather than earlier sessions
	The HAcs demonstrates good discriminative responsiveness
	The HAcs has the merit of being based on relatively literal signs and consequently would inform us about the frequency of different types of helping alliance signs, information which could not come from the HAr method[2]
	The psychometric properties of the HAcs appear to be somewhat better than those of the HAr[2]
	Provides a precise location for the location scored[1]
Weaknesses	Time-consuming to use
	Clinical experience is essential in order to use the HAcs
	Because the use of the instruments has been limited to transcripts, only the verbal channel of communication has been tapped
Areas for further research	More research is needed on the type 1 and type 2 groupings of scoring categories in the HAcs manual[2]

Primary reference

1. Alexander LB, Luborsky L. The Penn Helping Alliance Scales. In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: A research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford Press; 1986. pp. 325–66.
2. Luborsky L. A pattern-setting therapeutic alliance study revisited. *Psychother Res* 2000;**10**:17–29.
3. Luborsky L, Crits-Christoph P, Alexander LB, Margolis M, Cohen M. Two helping alliance methods for predicting outcome of psychotherapy. *J Nerv Ment Dis* 1983;**171**:480–91.
4. Luborsky L, Crits-Christoph P, Mintz J, Auerbach A. *Who will benefit from psychotherapy: predicting therapeutic outcomes*. New York: Basic Books; 1988.

Secondary references

None

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

H4
Hill Client Verbal Response Category System (HCVRCS)

General details	
Author	Hill C
Language	English
Country of publication/development	USA
Publication date	1981
Publisher	NA
Purpose and overview	
Study of response modes of clients. Response modes describe the client's style of involvement in the interaction and predict the ability to participate in a verbal therapy interaction[1]	
Theoretical orientation	Designed to be pan-theoretical.[1] The response modes are general to all situations.[1] The system seems to focus on most behaviours valued in client-centred or psychodynamic therapies, such as experiencing or insight, and to disregard behaviours valued more in cognitive or behavioural therapies, such as cognitive or behavioural exploration[2]
Population details	See below
Perspective	Trained independent observers
Measure used by	Can be used for training, practice and research[1]
Other versions	Hill Counselor Verbal Response Category System
Notes	*Raters:* Trained judges[1] Either upper-level undergraduate or graduate students are suitable, with preference given to people who have had some type of helping skills training[1] *Clients:* No details
Areas of therapist–patient interaction addressed: Map	
Individual differences: level of functioning	
Framework: convergent; reciprocal	
Patient engagement: intentions; motivation; expectation/preferences	
Nonverbal communication: silence	
Outcomes: changing view of self with others; working alliance (goals)	
Inferred from information provided below	
Dimensions	
Simple response	The nine categories are mutually exclusive
Request	A short and limited phrase (typically one or two words) which is usually of three types: (1) indicates agreement, acknowledgement, understanding or approval of what the counsellor has said; (2) indicates disagreement or disapproval with what the counsellor has said; or (3) responds briefly to a counsellor's question with specific information or facts. Generally, responses in this category do not indicate feelings, description or exploration of the problem
Description	An attempt to obtain information or advice or to place the burden or responsibility for solution of the problem on the counsellor

continued

	Discusses history, events or incidents related to the problem in a storytelling or narrative style. The person seems more interested in describing what happened than in communicating affective responses, understanding, or resolving the problem
Experiencing	Affectively explores feeling, behaviours or reactions about self or problems, but does not convey an understanding of causality. It may indicate a growing awareness of self or problems, but does not convey an understanding of causality. It may indicate a growing awareness of behaviours or problems without necessarily understanding why they have occurred, but does not refer to feelings towards counsellor/counselling situation
Exploration of client–counsellor relationship	Indicates feelings, reactions, attitudes or behaviours related to the counsellor or the counselling situation, but does not refer to feelings that are not directed towards the counsellor
Insight	Indicates that a client understands or is able to see themes, patterns or causal relationships in his or her behaviour or personality, or in another's behaviour or personality, and often has an 'a-ha' quality
Discussion of plans	Refers to action-orientated plans, decisions, future goals and possible outcomes of plans
Silence	A pause of 5 seconds (4 seconds is close enough) is considered the client's pause if it occurs between the counsellor's statement and the client's statement, within the counsellor's statement or immediately after a client's simple response
Other (unrelated to client problems)	Statements that are unrelated to the client's problem, such as small talk or salutations, comments about weather or events, or any statements that do not seem to fit into other categories owing to difficulties in transcription comprehensibility or incompleteness
Reliability	
Adequate inter-judge reliability has been demonstrated for the HCVRCS	
Split-half	NA
Internal consistency	NA
Inter-rater	Average kappas from the three studies that have used the client system are quite high (0.71, 0.77 and 0.92)[1]
	Hill *et al*. (1981) found high inter-judge agreement using this system (mean kappa = 0.92)[2]
Test–retest	No details
Validity	
No figures have been provided in the area of validity and so the conclusions made in refs 1 and 2 are questionable	
Face	Experts from several orientations who were used to establish face validity indicated that the HCVRCS covered the range of behaviours they would expect to occur within sessions[1]
Content	The HCVRCS was based on existing category systems, thus assuring a type of content validity[1]
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	The system found predicted relationships with counsellor interventions[2]
Convergent	No details
Discriminant	No details
Factor structure	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	Differences were found between the first and middle versus the final third of sessions. The probability of client description following counsellor minimal encourager decreased (75% vs 56%), whereas the probability of client experiencing increased (20% vs 38%). The probability of client simple response following counsellor information increased (60% vs 86%), whereas description increased (26% vs 4%). The probability of counsellor information following client simple response increased (35% vs 68%)
Acceptability	
Number of items	Nine categories
Administration method	Rating scale
Time taken to complete	Prospective judges should be made aware of the necessary attentiveness to detail and of the tediousness of the task[1]
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1981, Marathon Consulting and Press
Web or scanning options	No details
Training details	Trained judges are required. Judges in ref. 1 have always continued training until at least two of the three judges agree on 75–80% of all categorisations, which usually requires about 20 hours
Administration/process details	Transcripts of client audiotapes.[4] The operational principles are the same as for the Hill Counselor Verbal Response Category System: three trained judges are required to assign each response independently to one and only one category. Before categorising, typed transcripts of sessions must be divided into response units (essentially grammatical sentences) by trained unitisers[1]
Support from measure developers	Standardised training materials and techniques are available in a manual (Hill *et al*. *Manual for counselor and client verbal response category systems*, Columbus, OH: Marathon Consulting and Press; 1981) Sample transcript provided in ref. 1
FAQ facility	No details
Precision	
Scale type	Nominal, binary. The categories are pan-theoretical, they exhaust the range of possible behaviours at this level of analysis and cover behaviours observed in all theoretical orientations[1] A second type of judgement that would be helpful in summarising the data is the judges' determination of the 'predominant' or most impactful response within each speaking turn for both counsellor and client[1]
Normative data	Available in ref. 1
Notes	
The Client Behaviour System (CBS) was created to correct the deficiencies (see below) of the HCVRCS[2]	

continued

Résumé	
Strengths	Adequate inter-judge reliability has been demonstrated for the Hill Client Verbal Response Category System (HCVRCS)
	Any new researcher can easily use the systems using the manual[1]
Weaknesses	Most of the responses fell into the description (54%) and simple responses (25%) categories, which resulted in a restricted characterisation of the therapy process[2]
	There are no categories for describing client resistance, while research suggests that this is an important behaviour that needs to be included in a comprehensive measure of client verbal behaviour[2]
	The system includes categories that are at different conceptual levels; but from a methodological standpoint, measures should assess behaviours at the same level of abstraction (Greenberg and Pinsof, 1986)[2]
	The HCVRCS uses transcripts to code client behaviours, thus ignoring non-verbal and paralinguistic cues that would be available from the use of videotapes[2]
	The HCVRS relies on unit (sentence) judgements of client behaviours, which present a molecular analysis of what the client is saying. However, counsellors probably respond to the most predominant or central aspect of what the client communicates in an entire speaking turn[2]
	Requires attention to detail, and can be tedious[1]
	Extensive training required[1]
Areas for further research	Further testing of psychometric properties of the HCVRCS
	The addition of tapes would probably enhance reliability[1]
	No effort has been made as yet to do quality ratings on the response modes. For example, a good interpretation would lead to a different client response from a bad interpretation[1]
	The events leading up to a specific response have not been established[1]
	Future researchers may choose to subdivide simple responses into agreement vs disagreement[1]

Primary references

1. Hill CE. An overview of the Hill counselor and client verbal response modes category systems, In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford; 1986. pp. 131–59.
2. Hill CE, Corbett MM, Kanitz B, Rios P, Lightsey R, Gomez M. Client behavior in counseling and therapy sessions: development of a pantheoretical measure. *J Counsel Psychol* 1992;**39**:539–49.

Secondary references

3. Hill CE, Corbett MM, Kanitz B, Rios P, Lightsey R, Gomez M. Client behavior in counseling and therapy sessions: development of a pantheoretical measure, In Hill CE, editor. *Helping skills: the empirical foundation*. Washington, DC: American Psychological Association; 2001. pp. 21–40.
4. Meier A, Boivin M. Client Verbal Response Category System: preliminary data. *J Consult Clin Psychol* 1986;**54**:877–9.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

H5
Hill Interaction Matrix – Form G (HIM-G)

General details	
Author	Hill WF
Language	English
Country of publication/development	USA
Publication date	1975
Publisher	NA
Purpose and overview	
The Hill Interaction Matrix – Form G (HIM-G) is one of a family of four Hill Interaction Matrix (HIM) methods, which all have the same underlying concept, but varying formats	
The HIM is a behavioural rating system designed to measure the therapeutic qualities of group participant interactions. The HIM articulates an explicit value system for what is deemed to be therapeutic. The matrix is a weighted scoring system in which categories of behaviours are rank ordered on three therapeutic values: interpersonal threat, member centredness and therapist–patient role	
HIM-G was developed from the Hill Interaction Matrix – Statement by Statement (HIM-SS) in order to widen the application of the method, and requires less training to score. The matrix has been widely used to categorise group composition, leadership style and the status of group interaction	
Theoretical orientation	The measure was designed to be pan-theoretical and is based on observations from a wide variety of theoretical orientations
Population details	Clinical and non-clinical adults in group therapy
Perspective	Independent observer, group leader or group member
Measure used by	Researchers, counsellors, therapists
Other versions	HIM-SS, HIM-A, HIM-B
Notes	
Areas of therapist–patient interaction addressed: Map	
Roles	
Framework	
Patient engagement: motivation; commitment; intentions	
Threats to the relationship: hostility/anger; confrontations	
Outcomes: working alliance; goals; modification of working models	
Therapist-patient interaction information derived from the dimension descriptions	
Dimensions	
Content	The dimension relates to the content of groups' conversations, *what* they talk about, and has four categories: Relationships – IV (talking about group members' relationships and reactions to each other in the here and now), with four items (two each for leaders and members) Personal – III (talking about a member's problem in a historical sense), with four items (two each for leaders and members) Group – II (talking about the group itself), with four items (two each for leaders and members) Topic – I (talking about topics external to the group), with four items (two each for leaders and members)

continued

Work	The work dimension focuses on the *how* of the dialogue and categorises all interactions into five styles across two categories (pre-work, where members are not actively seeking self-understanding; and work, where a member is seeking self-understanding) *Pre-work:* Responsiveness (A): two statements (equally applicable to leaders and members) that probe to invite a member to interact, or the minimal response of a member who is unaccustomed to reacting interpersonally Convention (B): four statements or questions (two each for leaders and members) typical of an informal gathering Assertive (C): four statements or questions (two each for leaders and members) presented in an argumentative or hostile manner that suggest that the speaker cannot be influenced on the topic *Work:* Speculative (D): four statements (two each for leaders and members) representing the exchange of opinion and information to gain knowledge or clarification Confrontive (E): four statements (two each for leaders and members) representing an exchange that forces the members to come to terms with the essence of an idea experientially so that they can test it against their own experience
Reliability	
The picture regarding the reliability of the HIM-G is not clear. Adequate inter-rater reliability is reported in two studies[5,6] and a review by Hill (the author of the measure), which also refers to the HIM-G as replacing the HIM-SS.[2] Yet a reported later personal communication by Hill discusses problems with inter-rater reliability and interpreting internal consistency findings (e.g. see ref. 4), and suggests that HIM-G be used for cursory examination only, as the psychometric properties of the HIM-SS are superior[2]	
Split-half	No details
Internal consistency	96 pairwise correlations were for a modified HIM-G, both with zero values included and excluded with the following significance levels: 68 correlations significant at 0.01 and five significant at 0.05 (zeros included) Five were perfect, 21 significant at 0.01 and two significant at 0.05 (zeros excluded)[4]
Inter-rater	The quadrant scores from a modified HIM-G by two raters were correlated with the following results (*p* not reported): Topic-centred work, $r = 0.97$[5] Member-centred work, $r = 0.94$[5] Topic-centred pre-work, $r = 0.91$[5] Member-centred pre-work, $r = 0.89$[5] Using a modified HIM-G, inter-rater reliability coefficients ($n = 10$, *p* not reported) ranged from: 0.77 to 0.90 (Ebel, 1951; see ref. 6) 0.60 to 0.81 (Pearson product-moment)[6] Spearman's rho yields coefficients of 0.80[2]
Test–retest	NA
Validity	
The HIM-G has face and content validity as it was developed from the HIM-SS A modified HIM-G demonstrated adequate convergent validity with the Bonney scale (see ref. 6) when measured by a contingency coefficient and Pearson product-moment correlation, and χ^2 tests showed a significant association between the two.[6] Assessments of convergent validity of a modified HIM-G with Truax and Carkhuff's facilitative conditions (see ref. 5) suggested that the two measures enhance and complement each other rather than converge[5]	
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Face	See Content validity
Content	The HIM-G was developed from the HIM-SS
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	A modified HIM-G was assessed for convergent validity with the Bonney scale with the following results: χ^2 tests indicated a significant association at the 5% level[6] Contingency coefficient = 0.55[6] Pearson product-moment coefficient = 0.56, $p < 0.01$[6] Convergent validity of a modified HIM-G with Truax and Carkhuff's facilitative conditions was assessed with factor analyses. The results suggest that, while sharing properties that tap therapeutic interaction, the two measures enhance and complement each other rather than converge[5]
Discriminant	No details
Factor structure	Factor analyses of the HIM-G together with Truax and Carkhuff's facilitative conditions were conducted to assess the HIM-G's convergent validity (see Convergent validity and ref. 5 for further details)
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	72 statement items, 68 of which relate to the dimensions with four being non-specific. Many studies, including refs 5 and 6, have modified the matrix to 64 items by excluding the four Responsiveness statements and the four non-specific statements
Administration method	Rating scale
Time taken to complete	20 minutes
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1975, William Fawcett Hill
Web or scanning options	No details
Training details	Training is required, although scoring is far simpler than with the HIM-SS
Administration/process details	The scale is completed after viewing a group session, listening to an audiotape recording or reading a typed manuscript
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Likert-type each statement is rated on a six-point scale (0–5)
Normative data	Normative data for 50 diverse psychotherapy groups are reported in ref. 3

continued

Notes

Research studies employing the HIM-G include:

A comparison of the effects of facilitator-directed (FD) and self-directed (SD) personal growth group treatments on group member interaction[7]

Studies that evaluated an approach for treating acutely psychotic schizophrenic inpatients and outpatients in group settings[9,10]

Résumé

Strengths	Adequate inter-rater reliability is reported in two studies[5,6] and a review by Hill (the author of the measure). A modified HIM-G demonstrated adequate convergent validity with the Bonney scale when assessed with contingency coefficient and Pearson product-moment coefficient and χ^2 tests indicated a significant association at the 5% level.[6] The validity and reliability of the HIM-G are not clear (see also Weaknesses)
Weaknesses	There is some contradictory evidence regarding the reliability and validity of the HIM-G. Adequate inter-rater reliability is reported in two studies[5,6] and a review by Hill (the author of the measure), which also refers to the HIM-G as replacing the HIM-SS.[2] Yet a reported later personal communication by Hill discusses problems with inter-rater reliability and interpreting internal consistency findings (e.g. see ref. 4) and suggests that HIM-G be used for cursory examination only, as the psychometric properties of the HIM-SS are superior[2]
Areas for further research	Clarification of the measure's reliability and its value relative to HIM-SS

Primary references

1. Fuhriman A, Burlingame GM. The Hill Interaction Matrix: therapy through dialogue. In Beck AP, Lewis CM, editors. *The process of group psychotherapy: systems for analyzing change*. Washington, DC: American Psychological Association; 2000. pp. 135–74.
2. Hill WF. The Hill Interaction Matrix. *Person Guid J* 1971;**49**:619–23.
3. Hill WF. *Hill Interaction Matrix*. Los Angeles, CA: University of Southern California, Youth Study Center Press; 1965.
4. Powell ER. HIM Correlational Study. *Small Group Behav* 1977;**8**:369–80.
5. Roe JE, Edwards KJ. Relationship of two process measurement systems for group therapy. *J Consult Clin Psychol* 1978;**46**:1545–6.
6. Sisson CJ, Sisson PJ, Gazda GM. Extended group counseling with psychiatry residents: HIM and the Bonney Scale compared. *Small Group Behav* 1977;**8**:351–60.

Secondary references

7. Conyne RK, Rapin LS. A HIM-G interaction process analysis study of facilitator- and self-directed groups. *Small Group Behav* 1977;**8**:333–40.
8. Hill WF. Hill Interaction Matrix (HIM). The conceptual framework derived rating scales and an updated bibliography. *Small Group Behav* 1977;**8**:251–68.
9. Kanas N, Barr MA, Dossick S. The homogeneous schizophrenic inpatient group: an evaluation using the Hill Interaction Matrix. *Small Group Behav* 1985;**16**:397–409.
10. Kanas N, Smith AJ. Schizophrenic group process: a comparison and replication using the HIM-G. *Group* 1990;**14**:246–52.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

H6
Hill Interaction Matrix – Statement by Statement (HIM-SS)

General details	
Author	Hill WF
Language	English
Country of publication/development	USA
Publication date	1965
Publisher	NA
Purpose and overview	
HIM-SS is often referred to as the Hill Interaction Matrix (HIM) in the literature. However, the HIM is a family of four measures, with HIM-SS being one of them	
The HIM-SS is a behavioural rating system designed to measure the therapeutic qualities of group participant interactions. The HIM-SS articulates an explicit value system for what is deemed to be therapeutic. The HIM-SS is a weighted scoring system in which categories of behaviours are rank ordered on three therapeutic values: interpersonal threat, member centredness and therapist–patient role	
The HIM-SS has been used in research, group therapy and the training of group therapists	
Theoretical orientation	The measure was designed to be pan-theoretical and is based on observations from a wide variety of theoretical orientations
Population details	Children, adolescents and adults in group therapy
Perspective	Independent rater (usually a therapist)
Measure used by	Researchers, therapists and counsellors
Other versions	HIM-A, HIM-B, HIM-G
Notes	The development of the HIM-SS involved clients of the following therapies: group analytic, neuropsychoanalytic, pure psychoanalytic, non-directive, didactic, rational and guided group interaction[2,3]
Areas of therapist–patient interaction addressed: Map	
Roles	
Framework	
Patient engagement: motivation; commitment; intentions	
Threats to the relationship: hostility/anger; confrontations	
Outcomes: working alliance; goals; modification of working models	
The therapist-patient interaction information is derived from the items	
Dimensions	
Content	The dimension relates to the content of groups' conversations, *what* they talk about, and has four categories: Relationships – IV (talking about group members' relationships and reactions to each other in the here and now) Personal – III (talking about a member's problem in a historical sense) Group – II (talking about the group itself) Topic – I (talking about topics external to the group)

continued

Work	The work dimension focuses on the *how* of the dialogue and categorises all interactions into five styles across two categories (pre-work, where members are not actively seeking self-understanding; and work, where a member is seeking self-understanding) *Pre-work:* Responsiveness (A): a probe to invite a member to interact, or the minimal response of a member who is unaccustomed to reacting interpersonally Convention (B): statements or questions typical of an informal gathering Assertive (C): statements or questions presented in an argumentative or hostile manner that suggest that the speaker cannot be influenced on the topic *Work:* Speculative (D): exchange of opinion and information to gain knowledge or clarification Confrontive (E): exchange that forces the members to come to terms with the essence of an idea experientially so that they can test it against their own experience
Reliability	
The HIM-SS has been consistently reported as having adequate inter-rater reliability	
Split-half	No details
Internal consistency	No details
Inter-rater	The average percentage agreement across the 16 cells among three judges was 70%[2,3] Pearson product-moment correlation for three judges was 0.76[2,3] Spearman's rho was 0.90[2,3] Inter-rater reliability coefficients ranged from 0.89 to 0.94[1] Inter-rater reliability coefficients are typically reported from 0.72 to 0.92[2]
Test–retest	NA
Validity	
The HIM-SS was designed to be pan-theoretical and has face and content validity in that it was developed from observations of clients with a wide variety of diagnoses and therapists from a variety of orientations The HIM-SS Quadrant IV (scores with the greatest therapeutic value) has demonstrated adequate convergent validity with Truax and Carkhuff (see ref. 5) empathy and specificity scales, while correlations with respect and genuineness were not significant	
Face	See Content validity
Content	To achieve pan-theoretical application, the development of the HIM-SS involved clients of a wide variety of group therapies, e.g. group analytic, neuropsychoanalytic, pure psychoanalytic, non-directive, didactic, rational and guided group interaction therapies[2,3]
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	Quadrant IV of the HIM-SS was correlated (Pearson product-moment) with the Truax and Carkhuff facilitative conditions scales (see ref. 5), resulting in the following coefficients: HIM-SS with Empathy, 0.61 ($p < 0.01$)[5] HIM-SS with Specificity, 0.55 ($p < 0.01$)[5] HIM-SS with respect and genuineness, ns[5]
Discriminant	No details
Factor structure	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Responsiveness	
Discriminative (between individuals)	The HIM-SS has been responsive to differences in content and work style patterns between the following therapy groups: group analytic, neuropsychoanalytic, pure psychoanalytic, non-directive, didactic, rational and guided group interaction[2,3] Different patterns of matrix scores were found between an 'interaction' and an 'insight' group[3]
Evaluative (within individual across time)	The HIM-SS is responsive to developmental trends and the level of therapeutic quality across sessions[1]
Acceptability	
Number of items	20
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1965, William Fawcett Hill
Web or scanning options	No details
Training details	A high degree of training is required to score the HIM-SS[3]
Administration/process details	Recorded interactions are scored on a statement-by-statement basis
Support from measure developers	No details
FAQ facility	Frequently encountered scoring problems are discussed in the scoring manual[18]
Precision	
Scale type	Binary. The content and work items form the horizontal and vertical axis of the matrix. The matrix cells are weighted to represent different levels of therapeutic value given to interactions that are expected to facilitate therapeutic work and change Content items are valued from topic I (lowest) to relationship IV (highest) Work items are valued from responsive A (lowest) to confrontive D (highest)
Normative data	Normative data for 50 diverse psychotherapy groups are reported in ref. 3
Notes	
The matrix items are listed and described in ref. 2 There is an unpublished scoring manual[18] Research uses of the HIM-SS include: Comparison of nine analysis systems, including the HIM-SS[8,9] A study of reality vs client-centred models in group counselling[10] As study of client–counsellor interaction as a function of Whitehorn–Betz scores[11] An examination of the relative impact of facilitator-directed (FD) and self-directed (SD) personal growth group treatments on therapeutic verbal interaction[12] A comparison of the verbal interactions in counselling sessions between peer-led and counsellor-led adolescent groups[13,14] The effect of facilitator utterances on participant responses in a brief marriage enrichment group[15]	
	continued

A study to assess the effect of Kagan's Interpersonal Process Recall (IPR) videotape replay method on group work and compare it with a T-group method[16]

A comparison of the counselling and supervision processes[19]

A study of the effects of leadership style (cognitive vs experiential) on content and work styles of short-term therapy groups[20]

A study that used a combination of feedback and escape techniques with seven hospitalised clients to increase therapeutic interaction[21]

An investigation of the effects of an interpersonal growth contract and of leader experience on encounter group process and outcome[23]

A study of the effects of immediate feedback on the therapeutic content of group leaders' statements[24]

A study with children and adults of the therapeutic effectiveness of differentially targeted humorous remarks in group psychotherapy[25]

An investigation into the verbal behaviours of leaderless and therapist-led counselling groups[27,30]

An evaluative study of brief intervention models[28]

An examination of the impact of brief group psychotherapy on marital and sex roles[29]

An evaluation of changes in behaviour occurring as a result of a marathon group experience[31]

A study of the treatment process in saturation group therapy[32]

An evaluation of modelling and experiential procedures for self-disclosure training[33]

Résumé

Strengths	The HIM-SS was designed to be pan-theoretical and was developed from observations of clients with a wide variety of diagnoses and therapists from a variety of orientations. A scoring manual is available, which discusses frequently encountered scoring problems[18]
	The HIM-SS is consistently reported as having adequate inter-rater reliability.[1–3] The HIM-SS Quadrant IV (scores with the greatest therapeutic value) has demonstrated adequate convergent validity with Truax and Carkhuff (see ref. 5) empathy and specificity scales[5]
	The HIM-SS has shown responsiveness to differences in content and work style patterns between the following therapy groups: group analytic, neuropsychoanalytic, pure psychoanalytic, non-directive, didactic, rational and guided group interaction.[2,3] Different patterns of matrix scores were found between an 'interaction' and an 'insight' group.[3] The HIM-SS is responsive to developmental trends and the level of therapeutic quality across sessions[1]
	Normative data were compiled from transcripts of therapy sessions from over 1200 therapists. Normative data are available for over 50 diverse psychotherapy groups, which is a strength in itself. It also demonstrates the capacity of the measure to rate diverse groups[3]
Weaknesses	A high degree of training is required to score the HIM-SS[3]
Areas for further research	Further examination of psychometric properties. Only convergent validity with the Truax and Carkhuff scales has so far been addressed

Primary references

1. Fuhriman A, Burlingame GM. Measuring small group process: a methodological application of chaos theory. *Small Group Res* 1994;**25**:502–19.
2. Fuhriman A, Burlingame GM. The Hill Interaction Matrix: therapy through dialogue. In Beck A, Lewis CM, editors. *The process of group psychotherapy: systems for analyzing change*. Washington, DC: American Psychological Association; 2000.
3. Hill WF. *Hill Interaction Matrix*. Los Angeles, CA: University of Southern California, Youth Study Center Press; 1965.
4. Hill WF. The Hill Interaction Matrix. *Person Guid J* 1971;**49**:619–23.
5. Lambert MJ, DeJulio SS. Toward a validation of diverse measures of human interaction and counseling process. *Small Group Behav* 1977;**8**:393–5.
6. Lewis CM, Beck AP. A summary of the application of the systems of analysis to Group A, Session 3. In Beck AP, Lewis CM, editors. *The process of group psychotherapy: systems for analyzing change*. Washington, DC: American Psychological Association; 2000.

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Secondary references

7. Banet AG. The HIM as a group energy map. *Small Group Behav* 1977;**8**:396–7.
8. Beck AP, Lewis CM. Comparison of the systems of analysis: concepts and theory. In Beck AP, Lewis CM, editors. *The process of group psychotherapy: systems for analyzing change*. Washington, DC: American Psychological association; 2000.
9. Beck AP, Lewis CM, editors. *The process of group psychotherapy: systems for analyzing change*. Washington, DC: American Psychological Association; 2000.
10. Bigelow GS. Reality versus client-centered models in group counseling. *School Counsel* 1969;**16**:191–4.
11. Boyd RE. Whitehorn–Betz A-B score as an effector of client-counselor interaction. *J Counsel Psychol* 1970;**17**:279–83.
12. Conyne RK, Rapin LS. Facilitator- and self-directed groups: a statement-by-statement interaction study. *Small Group Behav* 1977;**8**:341–50.
13. Guttman MA. Verbal interactions of professional and peer led group counselling sessions. *Can J Counsel* 1989;**23**:103–12.
14. Guttman MJ. Verbal interactions of peer led group counselling. *Can J Counsel* 1987;**21**:49–58.
15. Hammond TM, Worthington EL. The effect of facilitator utterances on participant responses in a brief ACME-type marriage enrichment group. *Am J Fam Ther* 1985;**13**:39–49.
16. Hartson DJ, Kunce JT. Videotape replay and recall in group work. *J Counsel Psychol* 1973;**20**:437–41.
17. Hill WF. *Hill Interaction Matrix (HIM) scoring manual*. Los Angeles, CA: Youth Studies Center, University of Southern California; 1963.
18. Hill WF. Hill Interaction Matrix (HIM). The conceptual framework derived rating scales and an updated bibliography. *Small Group Behav* 1977;**8**:251–68.
19. Lambert MJ. Supervisory and counseling process: a comparative study. *Counsel Educ Supervis* 1974;**14**:54–60.
20. Lewis J, Mider PA. Effects of leadership style on content and work styles of short-term therapy groups. *J Counsel Psychol* 1973;**20**:137–41.
21. Lindberg FH, Morrill RS, Kilstrom DR. Group therapy with hospitalized patients: increasing therapeutic interaction using a feedback-escape technique. *Small Group Behav* 1974;**5**:486–94.
22. MacKenzie KR, Livesley WJ. Outcome and process measures in brief group psychotherapy. *Psychiatr Ann* 1986;**16**:715–20.
23. Magyar CW, Apostal RA. Interpersonal growth contracts and leader experience: their effects in encounter groups. *Small Group Behav* 1977;**8**:381–92.
24. Pattinson PR, Rardin MW, Lindberg FH. Effects of immediate feedback on the therapeutic content of group leaders' statements. *Small Group Behav* 1977;**8**:303–11.
25. Peterson JP, Pollio HR. Therapeutic effectiveness of differentially targeted humorous remarks in group psychotherapy. *Group* 1982;**6**:39–50.
26. Rae DS, Vathally ST, Manderscheid RW, Silbergeld S. Hill Interaction Matrix (HIM) scoring and analysis programs. *Behav Res Meth Instrum* 1976;**8**:520–1.
27. Seligman M, Sterne DM. Verbal behavior in therapist-led, leaderless, and alternating group psychotherapy sessions. *J Counsel Psychol* 1969;**16**:325–8.
28. Silbergeld S. Evaluation of brief intervention models by the Hill Interaction Matrix. *Small Group Behav* 1977;**8**:281–302.
29. Silbergeld S, Thune ES, Manderscheid RW. Marital role dynamics during brief group psychotherapy: assessment of verbal interactions. *J Clin Psychol* 1980;**36**:480–92.
30. Sterne DM, Seligman M. Further comparisons of verbal behavior in therapist-led, leaderless, and alternating group psychotherapy sessions. *J Counsel Psychol* 1971;**18**:472–7.
31. Uhlemann MR, Weigel RG. Behavior change outcomes of marathon group treatment. *Small Group Behav* 1977;**8**:269–80.
32. Vernalis FF, Holson DG, Shipper JC, Butler DC. The treatment process in saturation group therapy. *Psychother Theory Res Pract* 1972;**9**:135–8.

H7
Hill Counselor Verbal Response Category System (HCVRCS)

General details	
Author	Hill C
Language	English
Country of publication/development	USA
Publication date	1978
Publisher	No details
Purpose and overview	
To measure counsellor verbal behaviour with a focus on specific behaviours or skills, rather than conditions (such as empathy). The systems can be used for both counsellor training and practice	
Theoretical orientation	Derived from sociolinguistic theory, but applicable to diverse counselling orientations. Person-centred,[1] behavioural,[1] various/range,[3] psychotherapy,[14] personal construct psychotherapy and rationalist cognitive therapy[17]
Population details	Counsellors
Perspective	Trained judges rate counsellors' responses
Measure used by	Practitioners, training and research
Other versions	Hill Client Verbal Response Category System Hill Counselor Verbal Response Category System – Revised Also an eight-category counsellor system
Notes	*Clients:* Seven clients, male and female, with various problems, each engaged in one of seven types of therapy[1] Six male and six female clients seeking personal and social counselling[3] 8–12 year olds[12] *Practitioners:* Psychiatrist[1] Psychotherapist[1] Psychologist[3] Graduates[3,16] Undergraduates[16] Radio psychology talk show hosts[10]
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: openness; hope/encouragement; praise/affirmation Framework: controlling; structuring; directive; challenging Therapeutic techniques: reflection in action; counter-transference Non-verbal communication: silence Threats to relationship: confrontations; critical Inferred from categories and their full definitions as listed in ref. 3	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Dimensions	
	14 categories rather than dimensions
Minimal encourager	A short phrase that indicates simple agreement, acknowledgement or understanding
Approval–reassurance	Provides emotional support, approval or reinforcement
Information	Supplies information in the form of data, facts, resources, theory, and the like
Direct guidance	Consists of directions or advice that the counsellor suggests for the client, or for the client and counsellor together, either within or outside the counselling session
Closed question	A data-gathering enquiry that requests a one- or two-word answer, a yes or no, or a confirmation of the counsellor's previous statement
Open question	A probe requests a clarification of feelings or an exploration of the situation without purposefully limiting the nature of the response to a yes or no or a one- or two-word response
Restatement	A simple repeating or rephrasing of the client's statement(s) (not necessarily just the immediately preceding statements)
Reflection	A repeating or paraphrasing of the client's statement (not necessarily just the immediately preceding statements)
Non-verbal referent	Points out or enquires about aspects of the client's non-verbal behaviour, e.g. body posture, voice tone or level, facial expressions, gestures, and so on
Interpretation	This goes beyond what the client has overtly recognised
Confrontation	Contains two parts: the first part may be implied rather than stated and refers to some aspect of the client's message or behaviour; the second part usually begins with a 'but' and presents a discrepancy
Self-disclose	This usually begins with an 'I'; the counsellor shares his or her own personal experiences and feelings with the client
Silence	A pause of 5 seconds is considered the counsellor's pause if it occurs between a client's statement and a counsellor's statement or within the client's statement (except after a simple acceptance of the counsellor's statement)
Other	Statements that are unrelated to client's problems, such as small talk or salutations, disapproval or criticism of the client, or any other unclassifiable statements
Reliability	
Inter-rater agreement ranges from being partial to adequate across studies. Agreement is better in some categories than in others. No other areas of reliability have been addressed	
Split-half	NA
Internal consistency	NA
Inter-rater	Three judges each judged 3866 counsellor responses as to which category they belonged to. The kappas on all categorisations for all possible combinations of two judges were 0.79 (SE 0.01), 0.78 (SE 0.01) and 0.81 (SE 0.01), indicating high agreement across all judges. Disagreements were subjected to cluster analysis and there was overlap and confusion among certain categories, resulting in a modified 14-category system[3]
	Counsellor responses were each judged as to which of the 14 categories they belonged, by three judges. Kappas for the judgements on all categorisations for all possible combinations of the two judges were acceptable (0.68, 0.71 and 0.73). Agreement levels for individual categories indicated that in seven categories judges agreed on 73–100% of responses, and in four categories their agreement was between 15% and 35% of responses[5]

continued

	Average kappas from six studies are all reported to be quite high, ranging from 0.68 to 0.79. Agreement levels for individual categories, reported in three studies, indicate high reliability for six categories and low reliability for five categories. Regarding inter-rater reliability for individual categories, there were high agreement levels, indicating that high reliability was found for minimal encourager, silence, direct guidance, closed question, open question and non-verbal referent. Low agreement levels were found for approval–reassurance, restatement, reflection, interpretation and confrontation Several studies support the validity of the HCVRCS by showing that it describes the data in a way that makes good clinical sense[4] Three judges rated seven videotaped therapy sessions using the HCVRCS. Correlations were calculated between each pair of raters. The mean pairwise correlation for the system as a whole was 0.61 (range for the 14 categories was 0.48 –0.94) phi alpha [1]
Test–retest	No details
Validity	
Convergent validity for the HCVRCS ranged from partially adequate to adequate. Hill (1978) reports at least minimal face and content validity. There have been methodological problems in determining the predictive validity of the HCVRCS	
Face	The system was judged to have at least minimal face validity. (The fifth version of the HCVRCS was assessed using ten new judges. 83% of the examples had agreement between eight of the ten judges, and these 83% were retained in the final sixth version)[3]
Content	The system was judged to have at least minimal content validity. See above[3]
Criterion (a) concurrent	No details
Criterion (b) predictive	Efforts to link verbal response modes to both immediate and long-range outcomes have so far been hampered by methodological problems[4]
Construct	No details
Convergent	Intercorrelations between eight of the HCVRCS's categories and corresponding categories in five other systems (together) were calculated. The eight correlations ranged from 0.32 to 0.82. The HCVRCS showed adequate convergent validity for question, interpretation and confrontation, while its measure of self-disclosure appeared to differ from the others[1]
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	The HCVRCS discriminated between the verbal behaviors of Rogers, Ellis and Perls, showing that the therapists used the different categories of response in ways that correspond with their different theoretical positions[5] The HCVRCS discriminated between types of counsellors, e.g. career counsellors' responses were significantly correlated with Ellis ($r_7 = 0.86$), but not with Rogers ($r_7 = 0.21$)[6]
Evaluative (within individual across time)	Counsellor activity and counsellor verbal behaviour were analysed with a one-way repeated-measures ANOVA for differences across thirds of the sessions. The HCVRCS was responsive to counsellor changes in the latter third of the sessions[3] The HCVRCS detected changes in how each individual therapist used the categories over the course of the session[5]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Acceptability	
Number of items	14 categories
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1978, American Psychological Association
Web or scanning options	No details
Training details	Standardised training materials and techniques are available in a manual (Hill CE, *et al*. *Manual for counselor and client verbal response category systems*. Columbus, OH: Marathon Consulting and Press; 1981)
Administration/process details	CVRCS is composed of nine mutually exclusive, nominal and pan-theoretical therapist response modes organised into seven clusters: approval, directives (information, direct guidance), question (closed question, open question), paraphrase, interpretation, confrontation and self-disclosure. Three judges assign one response mode to each therapist response unit (a grammatical sentence) and to each speaking turn (predominant judge)
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Nominal, binary
Normative data	No details

Notes

The original 17 category counsellor system was modified to 14 owing to confusion in the ratings. The original system was based on 11 other systems from primarily client-centred and psychodynamic orientations

The initial version (developed from 11 existing systems) with 25 categories obtained low inter-rater agreement on categorisations from two judges. A second version with 24 categories again obtained low agreement. A third version with 25 categories, used with two practice therapy sessions, obtained high inter-judge agreement on categorisations (80% and 90% between the two judges)

Three counselling psychologists were given the (third version) 25-category definitions and asked to match examples of counsellor responses to the appropriate category. Low agreement led to a revision resulting in a 17-category fourth version. This version obtained agreement on 80% of the examples by two out of three new judges. A further revision was conducted with only the examples that obtained the highest agreement retained. This fifth version was assessed using ten new judges. 83% of the examples had agreement between eight of the ten judges, and these 83% were retained in the final (sixth) version[3]

Résumé	
Strengths	Inter-rater agreement ranges from being partial to adequate across studies
	Convergent validity for the HCVRCS ranges from partial to adequate. Hill (1978) reports at least minimal face and content validity
	The CVCRS demonstrates good responsiveness
Weaknesses	Psychometric information is limited
Areas for further research	Further testing of psychometric properties

continued

Primary references

1. Elliott R, Hill CE, Stiles WB, Friedlander ML, Mahrer AR, Margison FR. Primary therapist response modes: comparison of six rating systems. In Hill CE, editor. *Helping skills: the empirical foundation*. Washington, DC: American Psychological Association; 2001. pp. 9–19.
2. Friedlander ML. Counseling discourse as a speech event: revision and extension of the Hill Counselor Verbal Response Category System. *J Counsel Psychol* 1982;**29**:425–9.
3. Hill CE. Development of a counselor verbal response category. *J Counsel Psychol* 1978;**25**:461–8.
4. Hill CE. An overview of the Hill counselor and client verbal response modes category systems, In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford; 1986. pp. 131–59.
5. Hill CE, Thames TB, Rardin DR. Comparison of Rogers, Perls, and Ellis on the Hill Counselor Verbal Response Category System. *J Counsel Psychol* 1979;**26**:198–203.
6. Nagel DP, Hoffman MA, Hill CE. A comparison of verbal response modes used by master's-level career counsellors and other helpers. *J Counsel Dev* 1995;**74**:101–4.

Secondary references

7. Edwards HP, Boulet DB, Mahrer AJ, Chagnon GJ, Mook B. Carl Rogers during initial interviews: a moderate and consistent therapist. *J Counsel Psychol* 1982;**29**:14–18.
8. Fitzpatrick MR, Stalikas A, Iwakabe S. Examining counselor interventions and client progress in the context of the therapeutic alliance. *Psychotherapy* 2001;**38**:160–70.
9. Lee DY, Uhlemann MR. Comparison of verbal responses of Rogers, Shostrom, and Lazarus. *J Counsel Psychol* 1984;**31**:91–4.
10. Levy DA. Social support and the media: analysis of responses by radio psychology talk show hosts. *Prof Psychol Res Pract* 1989;**20**:73–8.
11. McLennan J, Twigg K, Bezant B. Therapist construct systems in use during psychotherapy interviews. *J Clin Psychol* 1993;**49**:543–50.
12. Mook B. Analyses of therapist variables in a series of psychotherapy sessions with two child clients. *J Clin Psychol* 1982;**38**:63–76.
13. Nehls N. Group therapy for people with borderline personality disorder: interventions associated with positive outcomes. *Issues Ment Health Nurs* 1992;**13**:255–69.
14. Stalikas A, Fitzpatrick M. Client good moments: an intensive analysis of a single session. *Can J Counsel* 1995;**29**:160–75.
15. Stalikas A, Fitzpatrick M. Relationships between counsellor interventions, client experiencing, and emotional expressiveness: an exploratory study. *Can J Counsel* 1996;**30**:262–71.
16. Thompson AP. Changes in counseling skills during graduate and undergraduate study. *J Counsel Psychol* 1986;**33**:65–72.
17. Winter DA, Watson S. Personal construct psychotherapy and the cognitive therapies: different in theory but can they be differentiated in practice? *J Construct Psychol* 1999;**12**:1–22.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

H8
Hill Counselor Verbal Response Category System – Revised (HCVRCS-R)

General details

Author	Friedlander ML
Language	English
Country of publication/development	USA
Publication date	1982
Publisher	No details

Purpose and overview

A revision of the Hill Counselor Verbal Response Category System, the HCVRCS-R is intended to tighten the system conceptually. Also, the addition of the variable 'degree of structure' (DS) is intended to represent the degree to which a counsellor's messages potentially structure subsequent client responses

Theoretical orientation	The system was developed from a sociolinguistic perspective, but for use across orientations. Psychodynamic,[1] behavioural,[2] person-centered[2]
Population details	Psychologist,[1] graduates,[1,4] students,[4] paraprofessional,[3] psychotherapist,[2] psychiatrist[2]
Perspective	Trained judges rate counsellors' responses
Measure used by	Therapists
Other versions	Hill Counselor Verbal Response Category System (HCVRCS)
Notes	*Clients:* One 32-year-old male with interpersonal problems – DSM classification of adjustment disorder with mixed anxiety depressed mood – engaged in 16 sessions of short-term dynamic psychotherapy (STDP)[1] Seven clients, male and female with various problems, each engaged in one of seven types of therapy[2] Undergraduates seeking help for personal and vocational problems[3] 23 undergraduates, 16 female, 16 white. Aged 18–43 years. All had T scores from 40 to 70 on the global severity index of the Hopkins Symptom Checklist-90–Revised. All had a self-reported problem with assertiveness in close personal relationships[4] *Practitioners:* Doctoral level counselling psychologist[1] 11 doctoral student counsellors[3] 23 doctoral students in counseling and clinical psychology. 13 female, 19 white. Aged 23–38 years. Experience ranged from 1 to 13 years. The majority adhered to behavioural techniques[4] *Raters:* Raters were two expert judges with prior minimal training to inter-rater kappa reliabilities of 0.83 and 0.85[3] 15 female rater students (14 white, 14 undergraduates), blind to study's hypotheses. Raters were trained for approx. 20 hours until an inter-judge reliability of 0.70 was reached[4]

continued

Areas of therapist–patient interaction addressed: Map	
Roles: expert/authority/leader	
Therapist engagement: hope/encouragement; praise/affirmation; openness; listening	
Therapeutic techniques: responsiveness/receptiveness/attunement; reflection in action	
Framework: challenging; structuring; directive	
Threats to the relationship: confrontations	
Based on category information	
Dimensions	
14 categories rather than dimensions:	
Minimal encourager	
Silence	
Approval–reassurance	
Information	
Direct guidance	
Closed question	
Open question	
Restatement	
Reflection	
Interpretation	
Confrontation	
Non-verbal referent	
Self-disclosure	
Other (statements unrelated to client problems)	
Reliability	
The inter-rater reliability ranges from partial to adequate for the HCVRCS-R. No other areas of reliability were addressed	
Split-half	NA
Internal consistency	NA
Inter-rater	Four judges categorised therapist responses. Kappa ranges on all possible combinations of any two judges were from 0.69 to 0.78. 69% of each judge's independent ratings reached a level of 75% agreement between judges[1]
	Three judges rated seven videotaped therapy sessions using the HCVRCS-R. Correlations were calculated between each pair of raters. The mean correlation, i.e. reliability, for the system as a whole was 0.59 (range for the nine categories was 0.53 to 0.94)[2]
	Near-perfect agreement on categories among three judges. No figures provided[3]
	Between the six judges for the nine categories, the inter-rater reliability was 0.77 for unit judgements and 0.91 for predominant judgements[4]
Test–retest	No details
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Validity	
The HCVRCS-R has at least minimal face and content validity. Evidence is mixed regarding the predictive validity of the HCVRCS-R. The convergent validity of the HCVRCS-R ranges from being partial to adequate	
Face	Minimal validity achieved. The 14-category HCVRCS was revised to the nine-category HCVRCS-R after an analysis of inter-rater discrepancies using transcripts from Hill *et al*. (1981). Following the manual for the CVRCS (Hill *et al*., 1981), three psychologists then matched counsellor response samples to the nine-category definitions with near-perfect agreement, thus achieving minimal face and content validity[3]
Content	Minimal validity achieved. See above[3]
Criterion (a) concurrent	No details
Criterion (b) predictive	The ability of five HCVRCS-R modes to predict outcome as measured by the Category System of Good Moments (CSCGM; Mahrer, 1988) was tested over three sessions (6, 9 and 15) using χ^2 tests. Only two response modes in one session predicted outcome (in session 9 higher confrontation and lower interpretation frequencies than expected were associated with the occurrence of a good moment). Hill *et al*. (1988) established predictive validity in that response modes were viewed as differentially helpful by both clients and therapists[1]
	Correlations between client activity and degree of structure (DS) tested the potential utility of DS as a predictor of client responses. DS and client activity were negatively correlated in the two sets of data tested (r = –0.33, and –0.32). However, it cannot be determined whether clients' passivity was a stimulus or response to DS[3]
Construct	Four studies have established construct validity with the system, in that it distinguishes therapists from different orientations in predictable ways[1]
Convergent	Intercorrelations between the HCVRCS-R categories and corresponding categories in five other systems (together) were calculated. The correlations ranged from 0.34 to 0.76. The HCVRCS-R showed good convergent validity for advisement and confrontation, but not self-disclosure[2]
	Two authors identified all possible psychotherapy Q-set (PQS) items that seemed similar to any of the CVRCS clusters. PQS items that corresponded to all seven CVRCS were identified. The approval cluster had an alpha of 0.91 and consisted of four items. The directives cluster had an alpha of 0.92 and consisted of three items. The paraphrase cluster had an alpha of 0.59 and consisted of two items. The interpretation cluster had an alpha of 0.69 and consisted of four items. None of the CVRCS clusters significantly correlated with the corresponding Q-set cluster[4]
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	Distinguishes therapists from different orientations in predictable ways[1]
Evaluative (within individual across time)	The CVRCS-R was responsive to changes in the DS present in an interview over time[3]
Acceptability	
Number of items	Nine
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

Feasibility	
Copyright	1982, American Psychological Association
Web or scanning options	No details
Training details	Raters receive training in order to attain inter-rater reliability
Administration/process details	Degree of structure (DS) is a derived variable, calculated after categorisation (high structure/moderate/low structure), representing a hypothetical continuum of the predictable effect of the counsellor's speech acts on subsequent client responses
Support from measure developers	Manual available from the author
FAQ facility	No details
Precision	
Scale type	Nominal, binary
Normative data	No details
Notes	
The HCVRCS-R is intended to simplify the HCVRCS conceptually by using only pragmatically coded, mutually exclusive intersubjective categories This paper discusses two conceptual and methodological inconsistencies of the HCVRCS that compromise its precision and utility. These are, first, that the classification rules require both classic and pragmatic coding strategies, which violates recommendations for conceptual rigour; and secondly, the rules for dividing the counsellors' discourse into meaningful units contain inconsistencies that affect the frequency counts and subsequent interpretations. The author has, therefore, refined the system with the intention of tightening it conceptually, and extended it, adding a 'degree of structure' variable. The revised version is the HCVRCS-R[3]	
Résumé	
Strengths	The inter-rater reliability ranges from partial to adequate for the HCVRCS-R. The HCVRCS-R has at least minimal face and content validity. The convergent validity of the HCVRCS-R ranges from being partial to adequate The HCVRCS-R demonstrates good responsiveness
Weaknesses	Few areas of reliability have been addressed
Areas for further research	Evidence is mixed regarding the predictive validity of the HCVRCS-R. Further psychometric validation is required
Primary references	
1. De Stefano J, Bernardelli A, Stalikas A, Iwakabe S. The relationship of therapist verbal response mode and client good moments in short term dynamic psychotherapy. *Can J Counsel* 2001;**35**:260–76. 2. Elliott R, Hill CE, Stiles WB, Friedlander ML, Mahrer AR, Margison FR. Primary therapist response modes: comparison of six rating systems. In Hill CE, editor. *Helping skills: the empirical foundation*. Washington; DC: American Psychological Association; 2001. pp. 9–19. 3. Friedlander ML. Counseling discourse as a speech event: revision and extension of the Hill Counselor Verbal Response Category System. *J Counsel Psychol* 1982;**28**:425–9. 4. Heaton KJ, Hill CE, Edwards LA. Comparing molecular and molar methods of judging therapist techniques. *Psychother Res* 1995;**5**:141–53.	
Secondary references	
None	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

11
Integrative Psychotherapy Alliance Scale (IPAS)

General details

Author	Pinsoff WM
Language	English
Country of publication/development	USA
Publication date	1986
Publisher	No details

Purpose and overview

Attends to the concept of the alliance in couple and family therapy, and brings an interpersonal and systemic perspective to bear on the concept of the alliance in individual psychotherapy[2]

The scales tap into the individual patient's perceptions of indicators of the alliance between the therapist and: (1) the patient (self–therapist); (2) the other members of the family or important members of the patient's interpersonal system (other–therapist); and (3) the family or interpersonal system of which the patient is a part (group–therapist)[2]

Developed according to Bordin's client and therapist relationship framework[1]

Theoretical orientation	Intergenerational family therapy or didactic family classes,[1] problem-centred therapy model.[2] Used for individual, couple and family therapy (separate scales)
Population details	Used with families, one member of whom has a heroin addiction
Perspective	Self-report
Measure used by	Practitioners, researchers
Other versions	Three versions: individual, family and couple
Notes	*Practitioners:* Therapists ranged in age from 24 to 56, with a mean of 28. Over 90% were Caucasian, and half were male[2] *Clients:* Families (identified patients and a family member) from three different methadone clinics in San Francisco, California[1] Ranged from 10 to 60 years of age, approx. two-thirds female, 80% Caucasian. The family cases presented with a variety of child problems, ranging from academic underachievement to adolescent conduct disorders. The marital cases presented primarily with relationship complaints ranging from chronic conflict to sexual dysfunction. The individual cases were unmarried adults with neurotic depression and anxiety disorders. No patients were psychotic[2]

Areas of therapist–patient interaction addressed: Map

Roles: expert/authority/leader

Individual differences: social support

Therapist engagement: genuineness

Patient engagement: attraction

Framework: convergent

Therapeutic techniques: transference/counter-transference

Outcomes: achieving a working relationship; working alliance (task/bond/goals)

Inferred from examples of items provided in ref. 2

continued

Dimensions	
Content	Refers to the thematic categories of the alliance: the *what* of the alliance. This has three categories: • Task: Bordin's (1979) concept of the tasks component of the alliance concerns the extent to which the methods and techniques of therapy are linked to "the patient's sense of his difficulties and his wish to change" (p. 254) • Bond: the bonds component refers to the quality of the human relationship between the therapist and the patient • Goals: the extent to which the therapist and patient agree on the goals of therapy is the relevant factor in the goals component In the current versions of the scales, the subdimensions of the content dimension differ in number of items. Tasks has the most, then bonds, and goals has the least
Interpersonal system	Refers to the human systems involved in the alliance: the *who* of the alliance. This has three categories: • self–therapist • other–therapist • group–therapist The interpersonal system dimension conceives of the alliance as a multi-systemic phenomenon that is manifested not only in the relationship between the reporting patient and the therapist, but also in the relationships between the therapists and other relevant members of the patient's interpersonal system
Reliability	
The internal consistency of the IPAS is adequate. However, the internal consistency of the dimensions ranges between inadequate and partial The IPAS demonstrates adequate test–retest reliability	
Split-half	No details
Internal consistency	Cronbach's alpha. The internal consistency of the total IPAS (range 0.79 to 0.87, average 0.84) was found to be adequate. The internal reliability coefficients for dimensions were: bonds 0.67, tasks 0.62, and goals 0.28. In previous studies in which this scale has been used, the internal reliability fluctuates between 0.79 and 0.83 with 17 and 35 participants, respectively (Pinsoff, 1983; Pinsoff & Catherall, 1986; Alvarez, 1991[1]
Inter-rater	NA
Test–retest	The IPAS Pearson correlation between the three administrations fluctuated between 0.73 and 0.80 at a significance level of 0.01. The correlation between third and sixth sessions was the highest (0.80). The correlation between the sixth and ninth sessions was the lowest (0.73)[1] Pearson correlations. Study 1 (using five-point scale): individual $r = 0.83$; couple $r = 0.84$; and family $r = 0.77$. All correlations exceeded the 0.005 level of significance. Study 2: all of the overall scores were high and significant at least at the 0.005 level: Individual $r = 0.72$; couple $r = 0.79$; and family $r = 0.83$. With one exception (goals on the family scale), all of the 18 subscale scores (six from each instrument) were significant at least at the 0.05 level. In a pilot study of four couple therapy cases, Gutterman (1984) found little variation in patients' scores on the couple Therapy Alliance Scale over the first eight sessions of therapy[2]
Validity	
The IPAS demonstrates adequate predictive validity with the Beck Depression Inventory (BDI) No figures are provided for convergent validity, and so although significant correlations have been found, we cannot rate their adequacy	
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Face	The three scales (individual, couple and family) are based on the same concept of the alliance, and derive from two theoretical dimensions: content (tasks, bonds and goals) and interpersonal system (self–therapist other–therapist and group–therapist). These two dimensions form a 3×3 matrix that was used to generate the items for each of the alliance scales[2]
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	At the time of writing, the authors are engaged in several studies testing the predictive validity of these scales. Preliminary results are encouraging[2] Although not statistically significant, early measures of IPAS were more predictive of BDI measure of depression than later measures. When this regression analysis was repeated with participants who received the Intergenerational family therapy (IFT) intervention, by the ninth session the alliance score explained 83% of the variability in outcome ($R^2 = 0.83$, $p < 0.01$). The BDI change score was significantly correlated with the total scores of the IPAS in the third administration ($r = 0.67$, $p < 0.05$), with the dimension of the tasks ($r = 0.69$, $p < 0.05$) and the dimension of bonds ($r = 0.64$, $p < 0.05$)[1]
Construct	No details
Convergent	Catherall (1984) found positive ($p < 0.05$) correlations between each of the overall alliance scale scores (individual scale $n = 28$; couple $n = 48$; family $n = 33$) and patient progress as measured by a therapist-report instrument developed by Storrow (1960) and modified for the conjoint contexts by Catherall (1984)[2]
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	Study 1: the distributions for each scale were highly skewed. Patients were very reluctant to say anything negative about their therapy or therapist. In study 2, the scale was enlarged from five to seven points in an attempt to reduce the skew. The distributions for the individual and couple scales levelled off, whereas the family scale distribution peaked and became more extreme. Moving from the five- to the seven-point scale did seem to increase the distribution of scores over the scale and will hopefully increase the likelihood that the scales will be able to predict negative and positive outcomes in individual, couple and family therapy[2] No significant differences were found among the IPAS mean score for the third, sixth and ninth psychotherapy sessions between treatment groups[1]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	Individual Therapy Alliance Scale has 25 items, and the Couple and Family Therapy Alliance Scales each consist of 29 items
Administration method	Rating scale Initially takes 5 minutes to fill out (2–4 minutes as patient become more familiar)[2]
Time taken to complete	Takes 5–10 minutes to complete[1]
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

Feasibility	
Copyright	1986, *Journal of Marital and Family Therapy*
Web or scanning options	No details
Training details	Manual available
Administration/process details	Administered at the end of a psychotherapy session, can be administered every session[2] Each of the instruments generates a minimum of seven scores or variables: an overall alliance score and a score on each of the three categories of the content (task, bond and goals) and interpersonal system (self–therapist, other–therapist and group–therapist) dimensions. The overall score is based on the mean rating of all of the items on the instrument. The six category scores are based on the mean of all of the items within each category (row or column of the matrix). The Couple and Family Therapy Alliance Scales, which are administered to all the family members present (above the age of 10), generate whole family and couple system scores (conjoint scores)
Support from measure developers	Manual. Copies of the most current, up-to-date versions of the Family Therapy Alliance Scale, the Couple Therapy Alliance Scale and the Individual Therapy Alliance Scale are available upon request from the authors
FAQ facility	No details
Precision	
Scale type	Likert. The items are presented in a seven-point scale that fluctuates from complete agreement (7) to complete disagreement (1)[1] The balanced scale and equal number of negative and positive items are intended to minimise the likelihood of experimental bias influencing test responses or expectations about treatment
Normative data	Mean scores provided
Notes	
Because different system members are present in individual, couple and family therapy, the other–therapist and group–therapist subdimensions had to be operationalised differently for each therapeutic context The authors are currently in the process of modifying and expanding their concept of the alliance as well as their attempts to measure it. The major theoretical modification involves the creation of a fourth interpersonal system subdimension (self–other) that assesses the status of the therapeutic alliance between the patient and other members of the patient system in regard to the three content subdimensions[2]	
Résumé	
Strengths	The internal consistency, test–retest reliability and predictive validity of the IPAS are adequate An instrument such as the IPAS may help with the goal of developing specific skills to create positive therapeutic alliances early in the treatment process[1] Quick to complete
Weaknesses	The results indicate that the internal consistency of the IPAS's goal dimension may be of questionable use since its reliability results only meet minimum empirical requirements for reliability[1] Owing to the inherent limitations of the rate–rerate test with a state-like variable such as the alliance, it is impossible to distinguish what part of the variance in scores derives from measurement error vs true and expected alliance variability[2]
Areas for further research	Future studies using other reliability and validity tests must be conducted to evaluate more adequately these aspects of the scales. The internal consistency of the scales, as well as the independence of the subdimensions, need to be evaluated[2]
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Primary references

1. Bernal G, Bonilla J, Alvarez MA, Greaux B. The psychotherapy alliance as a predictor of outcome: a preliminary study. *Rev Interam Psicol* 1993;**27**:229–38.
2. Pinsof WM, Catherall DR. The integrative psychotherapy alliance: family, couple and individual therapy scales. *J Marital Fam Ther* 1986;**12**:137–51.

Secondary references

None

I2
Intersession Experience Questionnaire (IEQ)

General details	
Authors	Orlinsky DE, Tarragona M
Language	English
Country of publication/development	USA
Publication date	1993
Publisher	NA
Purpose and overview	
The Intersession Experience Questionnaire (IEQ) examines patients' representations during the intervals between therapy sessions and is designed for repeated use over the course of treatment. The IEQ examines the functional value of patient representations, emphasising the states of mind in which they occur and the affective impact of their occurrence	
Theoretical orientation	Psychodynamic
Population details	Clinical adults
Perspective	Self-report
Measure used by	Researchers, psychotherapists and paraprofessionals
Other versions	A therapist form of the IEQ has been published. There are no primary article details regarding the psychometric properties of the form
Notes	The IEQ pilot samples were of clients of an outpatient psychiatric clinic (*n* = 279), a family treatment centre (*n* = 70) and private practice (*n* = 20). Therapy was individual, couples or family, with advanced trainee or experienced therapists of psychodynamic or integrative problem-centred orientation. Clients were aged between 20 and 39 years, male and female, primarily white, and the majority were college educated
Areas of therapist–patient interaction addressed: Map	
Derived from an item analysis	
Therapy context: type of therapy; boundaries; responsibilities	
Roles: attachment figure; expert/authority/leader	
Individual differences: attachment styles	
Patient engagement: motivation; attraction; commitment; intentions; expectation/preferences	
Framework: collaborative/participative/involving	
Threats to the relationship: hostility/anger	
Outcomes: compliance; working alliance: affective bond; goals; safety/secure base; expression of feelings	
Dimensions	
Recreating the therapeutic conversation	Defined as the patient's tendency to think about the therapist and therapy before and after sessions, between sessions and on the day
Evoking relief and remoralisation	Defined by positive-feeling items that clients reported feeling when thinking about their therapist or therapy
Evoking anxiety and frustration	Defined by negative-feeling items that clients reported feeling when thinking about their therapist and therapy
Preconscious or unconscious processing of therapeutic experience	The dimension is defined by the tendency of the patient to daydream, fantasise or actually remember having a dream about therapy or the therapist

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
During piloting of the IEQ, reliabilities of the four dimensions were assessed with Cronbach's alpha and demonstrated various degrees of adequacy at different stages of therapy. The coefficients for preconscious or unconscious processing of therapeutic experience, calculated from the fifth or sixth session, ranged from being partially supportive of internal consistency to adequate. The coefficients for the remaining three dimensions, calculated from the beginning of therapy, were all adequate (no details of probabilities are reported). Correlations for successive sessions were partial and adequate[1]	
Split-half	No details
Internal consistency	Alphas for preconscious or unconscious processing of therapeutic experience ranged from 0.67 to 0.74 Alphas for re-creating the therapeutic conversation, evoking relief and remoralisation and evoking anxiety and frustration ranged from 0.78 to 0.92 Correlations for successive sessions ranged from 0.57 to 0.81[1]
Inter-rater	NA
Test–retest	No details
Validity	
Pilot work with the IEQ has found two overall patterns of representations across samples receiving various forms of therapy[1] Factor analysis with varimax rotation of data from an outpatient sample ($n = 279$) revealed a four-factor structure. In an assessment of generality, the findings were compared to those of another factor analysis, using different procedures, with data from an aggregate of four different sample groups ($n = 90$). The results were partially replicated in that two of the factors were identified again, while another two factors were organised differently[1]	
Face	No details
Content	Seven of the items, which relate to the content and context of clients' representations, are taken from the Therapist Representation Inventory (Geller *et al*, 1981; see ref. 1)
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	The two patterns of representations found across samples are supportive-guiding and conflict-containing The four factors identified from the outpatient sample are recreating the therapeutic conversation; evoking relief and remoralisation; evoking anxiety; and frustration and preconscious or unconscious processing of therapeutic experience. The first two were also identified from the comparison sample group data
Responsiveness	
Discriminative (between individuals)	Relatively poorly functioning clients, as rated by themselves and their therapist, reported experiencing more representations of their therapy Clients scoring high on the dimension evoking relief and remoralisation were able to experience an emotional uplift between sessions by evoking representations of their therapy, while high scorers on evoking anxiety and frustration were not
Evaluative (within individual across time)	The dimension recreating the therapeutic conversation was particularly salient when patients were feeling distressed

continued

Acceptability	
Number of items	42
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1993, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	No details
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Likert-type. Each item enquires about the frequency of representations and the three response options are 'none', 'sometimes' and 'a lot'
Normative data	No details
Notes	
The scale items are listed by Orlinsky *et al*. (1993)[1]	
Copies of the IEQ (patient and therapist forms) may be obtained from David E. Orlinsky, Committee of Human Development, University of Chicago, Chicago, IL: 60637, USA	
Résumé	
Strengths	The clients in the pilot samples were drawn from private and public services in individual, couples and family therapy
	While the full picture regarding internal consistency is not clear, it appears promising as many correlation coefficients have been adequate
Weaknesses	The majority of clients in the pilot samples were mainly white and college educated
	While adequate internal consistency has been reported, it is not consistent across the dimensions or over the course of therapy sessions
Areas for further research	Further examination of psychometric properties with more diverse client groups, e.g. clarification of internal consistency and content validity and factor structure are the only validity criterion so far addressed
Primary reference	
1. Orlinsky DE, Geller JD, Tarragona M, Farber B. Patients' representations of psychotherapy: a new focus for psychodynamic research. *J Consult Clin Psychol* 1993;**61**:596–610.	
Secondary references	
None	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

M1
Maslach Burnout Inventory – Therapist and Client Versions (MBI-T and MBI-C)

General details	
Author	Linehan *et al*. (MBI-T and MBI-C) Maslach *et al*. (Original MBI)
Language	English and other languages
Country of publication/development	USA
Publication date	2000 (MBI-T and MBI-C) 1996 (Original MBI manual, 3rd edn)
Publisher	NA (original MBI published by Consulting Psychologists Press)

Purpose and overview

The Maslach Burnout Inventory was developed to measure burnout in individuals, e.g. in the human services, with educators and with workers in other occupations[2]

This summary focuses on two adaptations for use by clients and therapists in the field of psychotherapy, the MBI–Client (MBI-C) and MBI–Therapist (MBI-T)[1]

MBI-C: although the MBI is not intended to be used reciprocally for clients of professional care providers, Linehan *et al*. (2000) adapted the MBI questions to address emotional exhaustion from working with therapists, depersonalisation of therapists, as well as opportunities for contributing to the welfare of the therapist that clients might experience in psychotherapy (personal accomplishment). References to work or the job were changed to references to working with one's therapist, e.g. 'I feel emotionally drained from my work' was changed to 'I feel emotionally drained from working with my counselors/therapists.' It was thought that the MBI was an appropriate instrument to adapt to clients (MBI-C), for the following reasons:

- The client–therapist relationship is an intense and reciprocal one; the stresses of psychotherapeutic work are therefore expected to influence both therapists and clients
- Individuals with borderline personality disorder (BPD) are high utilisers of psychosocial services and, thus, are vulnerable to burnout from working with multiple mental health professionals
- It follows that assessment of client burnout is a reasonable measure of difficulty of the therapeutic relationship from the client's perspective

The MBI-T was developed to measure the burnout of therapists in particular. For the therapist version of MBI, references to work were changed to references to working with clients, e.g. 'I worry that this job is hardening me emotionally' was changed to 'I worry that working with my clients is hardening me emotionally'

Theoretical orientation	Brief psychotherapy[1] MBI-C used with clients taking part in trials comparing dialectical behaviour therapy to non-behavioural therapy Original MBI could be classed as occupational psychology
Population details	Clinical adults. The MBI-C study[1] was with clients with BPD. All participants reporting psychotherapy during the past year (n = 70). Women only, aged between 18 and 45, all met criteria for BPD. Further details of the sample, and exclusion criteria are given in ref. 1
Perspective	MBI-C: client self-report; MBI-T: therapist self-report
Measure used by	Psychotherapists. In ref. 1 30 therapists participated. All were considered to be experts in the treatment of suicidal BPD patients and had 10 or more years of psychotherapy experience. Included cognitive-behaviour therapists. Further details about the therapists are provided in ref. 1
Other versions	The original version is the MBI–Human Services Survey.[2] This has been adapted for use with educators (MBI–Educators Survey) and other occupations[4,5] (MBI–General Survey)
Notes	At pretreatment, clients responded to the items regarding all of the counsellors or therapists they had seen in the previous year.[1] At the 4-month point, both clients and therapists answered the questions in reference to their experiences with each other

continued

Areas of therapist–patient interaction addressed: Map	
Therapy context: influence; responsibilities (context)	
Therapist engagement: empathy/sensitivity; support/tolerance; respect (developing the relationship)	
Patient engagement: motivation; commitment (developing the relationship)	
Framework: reciprocal (maintaining the relationship)	
Therapeutic techniques: responsiveness/receptiveness/attunement; ruptures/repair (maintaining the relationship)	
Threats to the relationship: critical; hostility/anger (maintaining the relationship)	
Achieving a working relationship: working alliance (outcomes)	
Inferred from items	
Dimensions	
Three dimensions:[2]	
Emotional exhaustion (EE)	Nine items: higher score reflects more EE, e.g. I feel emotionally drained after working with my clients/therapist, I feel burned out from working with my clients/therapist
Depersonalisation (DP)	Five items: higher score reflects more DP, e.g. I've become callous towards my clients/therapist, I don't really care what happens to some of my clients/my therapist
Personal accomplishment (PA)	Eight items: lower score reflects diminished PA, e.g. I feel I'm positively influencing my clients'/therapists' lives, I have accomplished many worthwhile things working with my clients/therapists
Reliability	
The original MBI is an established measure that has been well validated.[2] The modified client and therapist versions cover the same issues, but in relation to the therapist or client, rather than the job or recipient of care	
Split-half	No details
Internal consistency	The correlations among the MBI-C factors were 0.37 ($p < 0.01$), 0.01, and 0.03 between EE and DP, EE and PA, and DP and PA[1]
	The correlations among MBI-T factors were 0.36 ($p < 0.05$), –0.13, and –0.31 between EE and DP, EE and PA, and DP and PA, respectively[1]
Inter-rater	NA
Test–retest	No details
Validity	
The original MBI is an established measure that has been well validated.[2] The modified client and therapist versions cover the same issues, but in relation to the therapist or client, rather than the job or recipient of care	
Face	The original MBI was developed from interviews/questionnaires with burnt-out workers, and from reviewing established scales[2]
	MBI-C: the constructs of emotional exhaustion and depersonalisation are face valid for a client's experience in therapy; the authors suggest that it may be harder to understand how personal accomplishment is related, but that it was included on the basis that the opportunity to contribute to a positive relationship with the therapist or to the welfare of the therapist would be important for the client[1]
Content	See above
Criterion (a) concurrent	No details
Criterion (b) predictive	Association between pretreatment client EE and 4-month therapist EE is $r = 0.27$, $p = 0.07$; between client DP and therapist EE is $r = 0.33$, $p = 0.02$, and between client DP and therapist DP is $r = 0.28$, $p = 0.06$

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Construct	No details
Convergent	Ref. 2 reports substantial details on the convergent validity of the original MBI, including correlations of MBI with independent behavioural ratings, job characteristics and other measures of outcome. Correlations range from 0.15 to 0.56.
Discriminant	Ref. 2 reports details on the discriminant validity of the original MBI It was predicted that job satisfaction was not exactly the same thing as burnout and would not be highly correlated. Correlations of EE, DP and PA with job satisfaction were –0.23, –0.22 and 0.17, respectively, ($p < 0.06$) None of the MBI subscales correlated significantly with Crowne–Marlowe Social Desirability Scale. Some trends for discriminative validity between burnout and depression, and burnout and occupational stress have been found and are reported in ref. 2.
Factor structure	MBI-C: principal axis factor analysis with oblique rotation was conducted on the client sample ($n = 70$) data.[1] Three factors were clearly indicated by the Scree plot and accounted for 49.0% of the variance. Three items loaded on different factors of the MBI-C than on the original MBI, loading on DP in the present study instead of EE. Excluding these three items, the range of factor loadings on Maslach's MBI vs the MBI-C was similar for both EE (0.54 to 0.74 and 0.47 to 0.84 for MBI and MBI-C, respectively) and for DP (0.41 to 0.67 and 0.53 to 0.69 for MBI and MBI-C, respectively). For the PA factor, the ranges were 0.43 to 0.59 and 0.35 to 0.86 for MBI and MBI-C, respectively[1] MBI-T: principal axis factoring with oblique rotation was conducted on the therapist sample ($n = 30$) data.[1] Three factors were indicated by the Scree plot and accounted for 54.5% of the variance. Three items failed to load on the same factors as Maslach's sample. As with the MBI-C, the range of factor loadings for the EE and DP factors of the MBI-T was comparable to Maslach's sample; for EE 0.54 to 0.74 and 0.46 to 0.92 for MBI and MBI-T, respectively. For the PA factor, the range was 0.43 to 0.59 and 0.45 to 0.73 for MBI and MBI-T, respectively; thus, somewhat stronger for the MBI-T than for the MBI[1] See ref. 2 for details of the factor structure of the original MBI
Responsiveness	
Discriminative (between individuals)	The authors examined the pretreatment relationship between the MBI-C and MBI-T and characteristics of clients and characteristics of therapists.[1] For both clients and therapists, age, educational level and amount of treatment received (for clients) or given to BPD clients (for therapists) were unrelated to both EE and DP. For clients, education level was correlated with PA ($r = 0.277$, $p < 0.05$) and for therapists age was correlated with PA ($r = 0.37$, $p < 0.05$)[1] The therapist sample showed less EE than the Maslach mental health sample and had a trend towards lower scores than the client sample. For DP, the client and therapists samples were similar to one another and both were lower than the Maslach sample. The three samples all differed on PA, with the client sample being the lowest, the Maslach mental health sample being the next, and the therapist sample having the highest PA
Evaluative (within individual across time)	No details
Acceptability	
Number of items	MBI-C: 22 items MBI-T: 22 items
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	The original MBI has several translations[2] including a Dutch translation (Schaufeli, 1993)
Access by ethnic minorities	No details

continued

Feasibility	
Copyright	MBI: 1996, Consulting Psychologists Press[2] MBI-C and MBI-T: 2000, *Cognitive and Behavioural Practice*
Web or scanning options	No details
Training details	No details
Administration/process details	Questionnaire
Support from measure developers	Manual available for the main/original MBI[2]
FAQ facility	No details
Precision	
Scale type	Items are responded to on a seven-point frequency scale (0 = never, to 6 = everyday)
Normative data	The main MBI has normative data for mental health and other professionals[2]
Résumé	
Strengths	The original MBI from which the client and therapist versions are developed is now recognised as a leading measure of burnout.[2] It has been translated into various languages and psychometric studies in different settings have continued to validate the three-dimensional structure of the measure[2]
Weaknesses	The MBI-C and MBI-T validation presented in ref. 1 is a only preliminary study of the adapted measures, and is with a specific client group (BPD). The personal accomplishment construct may be difficult for clients to relate to
Areas for further research	Further research and validation of the MBI-C and MBI-T in different therapeutic settings to help understand the role of burnout in the psychotherapy process and outcome

Primary references

1. Linehan MM, Cochran BN, Mar CM, Levensky ER, Comtois KA. Therapeutic burnout among borderline personality disordered clients and their therapists: development and evaluation of two adaptations of the Maslach Burnout Inventory. *Cogn Behav Pract* 2000;**7**:329–37.
2. Maslach CJ, Jackson SE, Leiter MP. *The Maslach Burnout Inventory manual*, 3rd ed. Palo Alto: Consulting Psychologists Press; 1996.

Secondary references

3. Schaufeli WB, Van Dierendonck D. The construct validity of two burnout measures. *J Organ Behav* 1993;**14**:631–47.
4. Leiter MP, Schaufeli WB. Consistency of the burnout construct across occupations. *Anxiety Stress Coping* 1996;**9**:229–43.
5. Schaufeli WB, Leiter MP, Maslach C, Jackson SE. The Maslach Burnout Inventory – General Survey. In Maslach C, Jackson SE, Leiter MP, editors. *Maslach Burnout Inventory Manual*, Vol. 3 Palo Alto, CA: Consulting Psychologists Press; 1996.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

M2
Missouri Identifying Transference Scale (MITS)

General details	
Author	Multon KD
Language	English
Country of publication/development	USA
Publication date	1996
Publisher	NA
Purpose and overview	
The MITS was designed as a measure that would aid the detection of transference reactions in therapy	
Theoretical orientation	Psychoanalytic/psychodynamic
Population details	Adult
Perspective	Therapist completed
Measure used by	Therapists with training in psychoanalysis/psychodynamic therapy
Other versions	None
Notes	In the study[1] piloting the MITS 16 clients (seven women and nine men) took part. All clients were Caucasian and education ranged from 2 years of college to a doctorate degree. Six therapists took part (four men, two women) and five were doctoral students and one held an advanced master's qualification
Areas of therapist–patient interaction addressed: Map	
Therapy context: confidentiality; boundaries	
Roles: friend/companion; attachment figure; confidant; good object; protector	
Individual differences: attachment styles; defensive style repression	
Patient engagement: expectation/preferences; attraction	
Therapeutic techniques: transference/counter-transference	
Threats to the relationship: intrusive; defensive; critical; hostility/anger; fear; sexual involvement; hidden agendas; resistance; confrontations; withdrawal	
Outcomes: emotional expression	
Inferred from items	
Dimensions	
Negative transference reaction (NTR)	25 items, e.g. resentment; passivity; mistrust
Positive transference reaction (PTR)	12 items, e.g. admiration; idealisation
Reliability	
Cronbach's alpha was used as a measure of internal consistency: both scales showed adequate internal reliability	
Split-half	No details
Internal consistency	Cronbach's alpha: NTR (0.96); PTR (0.88) Item total correlations: 24 of the 37 correlations in the 0.61 to 0.82 range
Inter-rater	NA
Test–retest	No details

continued

Validity	
Concurrent validity with Luborsky's Psychotherapy Check Sheet (PCS; Graff & Luborsky, 1977) was tested. The PTR showed partial concurrent validity and the NTR showed adequate concurrent validity	
Partial convergent validity was demonstrated with the Interpersonal Schema Questionnaire (ISQ; Safran and Hill, 1989)	
Factor analysis showed that the MITS consists of two factors, positive and negative transference accounting for 52.3% of the variance	
Face	Items were written and rewritten through an iterative process to describe overt behavioural indicators of the client's transferential reactions to the therapist
Content	No details
Criterion (a) concurrent	The PTR was correlated 0.31 ($p < 0.01$) with the one item amount of positive transference from the PCS and the NTR was correlated 0.53 ($p < 0.01$) with the one item amount of negative transference from the PCS. The NTR was also significantly negatively correlated at –0.39 ($p < 0.01$) with the one item amount of positive transference from the PCS[1]
Convergent	No details
Criterion (b) predictive	Other authors[2,3] have highlighted that MITS does not clearly define 'transference', so construct validity is difficult to measure
Construct	No details
Convergent	In a cross-validation study[1] the NTR was significantly and negatively correlated with the control (–0.39, $p < 0.05$) and sociability (–0.38, $p < 0.05$) scales of the ISQ
Discriminant	No details
Factor structure	The 37 items were subjected to an iterative principal components extraction with a two-factor oblique rotation to obtain final loadings for the 25 items on the first factor and the 12 items on the second factor. The two factors were labelled negative transference reaction (NTR) and positive transference-reaction (PTR). The NTR accounted for 42.3% of the variance, and the PTR an additional 10%[1]
Responsiveness	
Discriminative (between individuals)	Clients who viewed their mothers as more controlling, untrustworthy, less affiliative and less sociable showed more negative transference reactions (as measured by ISQ)
Evaluative (within individual over time)	No details
Acceptability	
Number of items	37
Administration method	Completed by therapist post-session
Time taken to complete	No details
Flesch reading age	NA
Translations	No details
Access by ethnic minorities	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Feasibility	
Copyright	1996, American Psychological Association
Web or scanning options	None
Training details	Training consisted of one 2-hour session in which clinical examples were presented and discussed relative to the MITS ratings given to them by the counsellors
Administration/process details	Therapists completed the MITS after every session and directions read as follows: 'During the previous session the client had the following unrealistic reactions'
Support from measure developers	None
FAQ facility	None
Precision	
Scale type	Likert five-point scale ranging from 1 (not evident) to 5 (very evident). The 37 items were randomly arranged in an effort to reduce the possibility of response bias
Normative data	See ref. 1
Notes	
Limitations of pilot study:[1] a relatively small number of clients (16) was used. A larger number of clients and counsellors would have allowed exploration of possible moderating variables such as gender of client/counsellor, experience level of counsellor and type of client problem Until the psychometric properties of the scale are further clarified, MITS is not recommended for clinical decision-making; however, it does appear to be useful for research purposes	
Résumé	
Strengths	Adequate internal consistency for both scales. Initial evidence of concurrent validity with the PCS. Does not rely on session transcripts/videos like traditional rating scales
Weaknesses	Measure developers do not provide a compact single statement of their operational definition of transference[2,3]
	Only therapist is used as transference rater, introducing substantial bias
Areas for further research	Validity of the MITS to be tested using independent raters
	Further general testing of psychometric properties in larger samples with ethnically diverse participants

Primary references

1. Multon KD, Patton MJ, Kivlighan DM Jr. Development of the Missouri Identifying Transference Scale. *J Counsel Psychol* 1996;**43**:243–52.
2. Mallinckrodt B. Capturing the subjective and other challenges in measuring transference: comment on Multon, Patton, and Kivlighan. *J Counsel Psychol* 1996;**43**:253–6.
3. Carter JA. Measuring transference: can we identify what we have not defined? *J Counsel Psychol* 1996;**43**:257–8.
4. Multon KD, Patton MJ, Kivlighan DM, Counselor recognition of transference reactions: reply to Mallinckrodt (1996) and Carter (1996). *J Counsel Psychol* 1996;**43**:259–60.

Secondary references

None

M3
Multicultural Counseling Inventory (MCI)

General details	
Author	Sodowsky GR
Language	English
Country of publication/development	USA
Publication date	1994
Publisher	NA
Purpose and overview	
A measure of competencies of any counsellor working with a minority of culturally diverse clients. There is an assumption of a degree of overlap between counselling competence in general and multicultural counselling competencies, so the measure includes items on general counselling skills as well	
Theoretical orientation	Counselling psychology
Population details	Adults
Perspective	Therapist
Measure used by	Counselling psychologists
Other versions	No details
Notes	
Areas of therapist–patient interaction addressed: Map	
Therapy context: type of therapy; values; responsibilities	
Roles: advocate; confidant; protector	
Therapist engagement: empathy/sensitivity; warmth; genuineness; respect; support/tolerance; openness; listening	
Inferred from dimension and sample item information	
Dimensions	
Skills	11 items. Items referring to success with retention of minority clients, recognition of and recovery from cultural mistakes, use of non-traditional methods of assessment, counsellor self-monitoring, tailoring structured vs unstructured therapy to the needs of minority clients
Awareness	Ten items. Items reflecting proactive multicultural sensitivity and responsiveness, extensive multicultural interactions and life experiences, broad-based cultural understanding, advocacy within institutions, enjoyment of multiculturalism and an increase in minority caseload
Knowledge	11 items. Items referring to culturally relevant case conceptualisation and treatment strategies, cultural information and multicultural counselling research
Relationship	Eight items. Items referring to the counsellor's interactional process with the minority client such as the counsellor's trustworthiness, comfort level, stereotypes of the minority client and world-view

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
Across studies, the total scale demonstrated adequate internal consistency using Cronbach's alpha estimates. The dimensions showed partial to adequate internal consistency. See studies for further details	
Split-half	NA
Internal consistency	Total scale 0.88; dimensions range 0.65 to 0.83[1,6] Total scale 0.86; dimensions range 0.67 to 0.81[2,6] Dimensions range 0.63 to 0.76[4] Total scale 0.86; dimensions range 0.67 to 0.76[5] Total scale 0.89; range 0.67 to 0.83[1]
Inter-rater	NA
Test–retest	No details
Validity	
Convergent validity with the Multicultural Counseling Awareness Scale – Revised: Form B (MCAS) was tested.[4] Adequate convergent validity was found between the MCI knowledge and the MCAS knowledge skills scales. However, there are high correlations between each of these scales and MCI awareness, suggesting that the correlation of the knowledge scales across the two instruments is not conclusive evidence of convergent validity of the MCI and MCAS. Convergent validity of the MCI and MCAS total scores was adequate in study.[5] However, the scale intercorrelations suggest that different constructs are being measured[5]	
Face	An extensive review of multicultural counselling literature was used to develop the items[6]
Content	Experts judged item clarity and content. High inter-rater agreement (75–100%) was demonstrated on the relationship of item content to the names given to the four subscales[6]
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	MCI awareness with MCAS awareness: $r = 0.16$ (ns) MCI knowledge with MCAS knowledge/skills: $r = 0.58$ ($p < 0.01$) (see ref. 4 for full list of correlations) MCI total with MCAS total: $r = 0.64$ ($p < 0.05$) (see ref. 5 for full list of scale correlations)
Discriminant	No details
Factor structure	The scale comprises four factors of skills, awareness, relationship and knowledge, which has been demonstrated to account for 36.1% of the variance[1,6] and 35.3 of the variance[2,6] Inter-subscale correlations have ranged from 0.19 to 0.54[4] and from 0.27 to 0.56[6]
Responsiveness	
Discriminative (between individuals)	Respondents who worked 50% or more in the multicultural area scored significantly higher on the multicultural awareness and multicultural counselling relationship factors than respondents whose counselling work consisted of less than 50% minority service[6] All four ethnic minority counsellor groups reported higher MCI total scores and subscale score than did white counsellors. Counsellors who endorsed more than one ethnic minority category (multiracial) had the highest scores overall on MCI total and four subscale scores[1]
Evaluative (within individual across time)	Counsellors were rated more highly on the MCI following a multicultural training course[5]

continued

Acceptability	
Number of items	40
Administration method	Self-report questionnaire
Time taken to complete	15–25 minutes
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	Questionnaire is designed for counsellors working with ethnic minorities or culturally diverse clients
Feasibility	
Copyright	1994, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	Respondents are asked to 'indicate how accurately each statement describes you when working in a multicultural counseling situation'
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Each item is rated on a four-point Likert scale, ranging from 1 (very inaccurate) to 7 very accurate. Greater scores indicate higher multicultural awareness
Normative data	Available in refs 1, 4 and 5
Résumé	
Strengths	Adequate internal consistency. Measure is responsive: discriminates between individuals on criteria relevant to multicultural counselling competencies
Weaknesses	Mixed evidence for convergent validity with the MCAS, owing to lack of validity evidence for such recently developed instruments
Areas for further research	Further convergent validity testing to provide understanding of the differences between the MCI and the MCAS
	Investigation of the effect of social desirability responses
	Research on use and application of the MCI in multicultural awareness training

Primary references

1. Bellini J. Correlates of multicultural counseling competencies of vocational rehabilitation counselors. *Rehabil Counsel Bull* 2002;**45**:66–75.
2. Boyle DP, Springer A. Toward a cultural competence measure for social work with specific populations. *J Ethnic Cult Divers Soc Work* 2001;**9**(3–4):53–71.
3. Ponterotto JG, Rieger BP, Barrett A, Sparks R. Assessing multicultural counseling competence: a review of instrumentation. *J Counsel Dev* 1994;**72**:316–22.
4. Pope-Davis DB, Dings JG. An empirical comparison of two self-report multicultural counseling competency inventories. *Measure Eval Counsel Dev* 1994;**27**:93–102.
5. Pope-Davis DB, Dings JG. The assessment of multicultural counseling competencies. In Ponterotto JG, Casas JM, Suzuki LA, Alexandra CM. editors. *Handbook of multicultural counseling*.Thousand Oaks, CA; 1995. pp. 287–311.
6. Sodowsky GR, Taffe RC, Gutkin TB, Wise SL. Development of the Multicultural Counseling Inventory: a self-report measure of multicultural competencies. *J Counsel Psychol* 1994;**41**:137–48.

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Secondary references

7. Pope-Davis DB, Ottavi TM. Examining the association between self-reported multicultural counseling competencies and demographic variables among counselors. *J Counsel Dev* 1994;**72**:651–4.
8. Robles-Pina RA, McPherson RH. The relationship between educational and demographic variables and supervisor's multicultural counseling competencies. *Clin Superv* 2001;**20**:67–79.
9. Rubin SE, Davis EL, Noe SR, Turner TN. Assessing the effects of continuing multicultural rehabilitation counseling education. *Rehabil Educ* 1996;**10**:115–26.
10. Sodowsky GR, Kuo-Jackson PY, Richardson MF, Corey AT. Correlates of self-reported multicultural competencies: counselor multicultural social desirability, race, social inadequacy, locus of control racial ideology, and multicultural training. *J Counsel Psychol* 1998;**45**:256–64.

O1
Octant Scale Impact Message Inventory (IMI-C)

General details	
Author	Keisler DJ
Language	English
Country of publication/development	USA
Publication date	1997
Publisher	NA
Purpose and overview	
To describe the impact messages or covert reactions (i.e. feelings, cognitions, behavioural tendencies) one person (the rater) may characteristically experience in the presence of another (the target). This inventory is a revision of the Impact Message Inventory (IMI), which was derived from literature on interpersonal behaviour, especially Lorr and colleagues	
Theoretical orientation	Interpersonal
Population details	Adults: clinical and non-clinical
Perspective	Counsellors, partners and/or significant others rate their own responses to the subject of the inventory
Measure used by	Researchers, psychiatrists, counsellors, clinicians
Other versions	Impact Message Inventory
Notes	*Clients:* 168 undergraduates, 45% male, 80 European American, 83% aged 18–21, participating for course credit[1]
Areas of therapist–patient interaction addressed: Map	
Therapy context: power/coercion	
Roles: friend/companion; expert/authority/leader	
Patient engagement: expectation/preferences	
Framework: complementary; controlling; structuring; directive; flexible/rigid	
Threats to the relationship: hostility/anger	
Inferred from octants' information	
Dimensions	
Each octant scale comprises seven items	
Octants	e.g. 'When I am with this person, he/she makes me feel … … bossed around
Dominant	
Hostile–dominant	… that I want to stay away from him/her
Hostile	… distant from him/her
Hostile–submissive	… that I should tell him/her not to be so nervous around me
Submissive	… in charge
Friendly–submissive	… that I could tell him/her anything and he/she would agree
Friendly	… appreciated by him/her
Friendly–dominant	… that I could relax and he/she'd take charge
Dimensions	
Control	
Affiliation	
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
Partial to adequate internal consistency has been demonstrated for the IMI-C	
Split-half	No details
Internal consistency	Coefficients for each octant or category were calculated across eight subsamples and the sample as a whole. Cronbach's alphas for each octant scale in the sample as a whole ranged from 0.69 to 0.89, and from 0.60 to 0.90 for the eight subsamples. Subsample 3 produced relatively low alphas for three octant scales (friendly–dominant $r = 0.29$, hostile–submissive $r = 0.48$ and friendly submissive $r = 0.41$)[2]
Inter-rater	No details
Test–retest	No details

Validity	
Partial convergent validity for both dimensions of the IMI-C has been demonstrated with the extraversion and agreeableness dimensions of the Neuroticism–Extroversion Openess Inventory (NEO) Personality Inventory – Revised [NEO-PI-R (Revised NEO Personality Inventory); Peer Rating Form]. Partial convergent validity of NEO-PI-R and IMI-C has been demonstrated. Findings regarding factor structure are mixed	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	Participants completed the IMI-C and NEO-PI-R (rates five personality domains: neuroticism, extraversion, openness to new experience, agreeableness and conscientiousness) rating scales As expected, extraversion and agreeableness showed the highest correlations with the underlying IMI-C dimensions (extraversion correlated with control and affiliation 0.32 and 0.36, respectively; agreeableness correlated with control and affiliation –0.57 and 0.36 respectively, all $p < 0.01$) Additionally, significant correlations were found between neuroticism and both IMI–Client dimensions (–0.27, –0.26, $p < 0.01$), while conscientiousness (0.30) and openness (0.28) both had significant positive correlations with affiliation, yet showed no strong relationship to control[1]
Discriminant	No details
Factor structure	In analysing the NEO-PI-R in relation to the IMI-C, the findings are surprising as they show that the IMIC interpersonal circle is sensitive to (at least) some of the components of the supposed 'non-interpersonal' factors of the NEO-PI-R. The researchers seem to interpret the unexpected findings as most likely due to problems with the NEO-PI-R, rather than the IMI-C[1] Circumplex structure was evaluated with three strategies: principal component analysis (PCA) and post hoc χ^2, multidimensional scale analysis (MSA) and confirmatory factor analysis (CFA). CFA revealed that goodness of fit to true circumplex was less than desirable for the fully constrained models. The CFA is considered a more conservative test than the previous two and the results suggest that the IMI-C falls short of 'true circumplex' status[2] To assess whether IMI-C primary axes (dominance and affiliation) are orthogonal, the angle of separation is taken. In a two-dimensional space the angle should be 90 degrees and is 89.7 degrees, indicating that the dominance and affiliation dimensions are more or less orthogonal. Vector lengths (expressed in scores from 0 to 1) represent the extent to which a scale is represented by the two interpersonal factors (dominance and affiliation). Vector lengths of the IMI-C octants ranged from 0.78 to 0.88, indicating strong relationships with the interpersonal factors. The Fisher (1983)/Fisher *et al*. (1985) method of calculating cosines to assess the discrepancies between predicted and actual locations of a set of scales was employed. There was 95.2% agreement between predicted and actual locations. These analyses demonstrate adequate circumplex properties[3]

continued

Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	56
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1997, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	Raters are asked to imagine that they are in the presence of the subject of the inventory (target). Each of the eight subscales contains seven items, all beginning with 'When I am with this person he/she makes me feel … (e.g. that I could relax)'. Raters indicate on a scale of 1 (very much so) to 4 (not at all) how accurately each item describes their reaction to the target[1,2]
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert. Four-point scale, indicating the accuracy with which the item describes their reactions to the target (1 = not at all, 2 = sometimes, 3 = quite often and 4 = very much so)
Normative data	No details
Résumé	
Strengths	Partial to adequate internal consistency has been demonstrated for the IMI-C. Partial convergent validity for both dimensions of the IMI-C has been demonstrated with the extraversion and agreeableness dimensions of the NEO-PI-R; Peer Rating Form. Partial convergent validity of NEO-PI-R and IMI-C has been demonstrated
Weaknesses	Little psychometric information available
Areas for further research	Other areas of reliability need to be addressed
	Future work could clarify the mixed findings regarding factor structure
Primary references	
1. Schmidt JA, Wagner CC, Kiesler DJ. Covert reactions to Big Five personality traits: the Impact Message Inventory and the NEO-PI-R. *Eur J Psychol Assess* 1999;**15**:221–32. 2. Schmidt JA, Wagner CC, Kiesler DJ. Psychometric and circumplex properties of the octant scale Impact Message Inventory (IMI-C): a structural evaluation. *J Counsel Psychol* 1999;**46**:325–34. 3. Kiesler DJ, Schmidt JA, Wagner CC. A circumplex inventory of impact messages: an operational bridge between emotion and interpersonal behavior. In Plutchik R, Conte HR, editors. *Circumplex models of personality and emotions*. Washington, DC: American Psychological Association; 1997. pp. 221–44.	
Secondary references	
None	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

P1
Patient Action Scale (PAS)

General details	
Authors	Hoyt MF, Marmar CR, Horowitz MJ, Alvarez WF
Language	English
Country of publication/development	USA
Publication date	1981
Publisher	NA
Purpose and overview	
The purpose of the Patient Action Scale is to assess specific patient actions during dynamic psychotherapy. The PAS assesses the occurrence of actions and their emphasis in relation to the overall action of the session. The scale was devised to assess patient actions that could be identified and repeated if found to be related to successful treatment outcomes	
Theoretical orientation	Psychodynamic
Population details	Clinical adults
Perspective	Independent rater
Measure used by	Researchers, psychiatrists, postdoctoral fellows, psychiatric social worker, advanced psychiatry residents and psychodynamically orientated therapists
Other versions	No details
Notes	In the initial assessment of the scale, the clients were neurotic-level outpatients seen for a stress response syndrome following a stressful life event (e.g. bereavement). The clients numbered 25, had a mean age of 36.1 years, 21 were female and they received brief time-limited dynamic psychotherapy The raters were all psychodynamically trained The purpose of this initial assessment was to assess the reliability of the PAS in dynamic short-term psychotherapy
Areas of therapist–patient interaction addressed: Map	
Derived from an item analysis Therapy context: type of therapy; responsibilities Roles: expert/authority/leader Individual differences: defensive style/repression Patient engagement: expectation/preferences Framework: collaborative/participative/involving Non-verbal communication: laughter/humour; silence Information inferred from full scale	
Dimensions	
No details	
Reliability	
In the initial assessment of the PAS, inter-rater reliabilities ranged from partial to adequate; test–retest reliabilities ranged from inadequate to adequate; and the median for both sets of tests was adequate[1] (probabilities not reported)	
Split-half	No details
Internal-consistency	No details
Inter-rater	Finn's *r* ranged from 0.45 to 0.95, with a median of 0.75
Test–retest	Finn's *r* ranged from 0.54 to 0.99, with a median of 0.87

continued

Validity	
The PAS has face and content validity in that it was derived from previous activity scales and revised on the basis of suggestions from supervisors and advanced psychiatry residents from the authors' psychotherapy group. Three of the initial 27 items were excluded from the scale because they were endorsed for less than 20% of sessions[1]	
In assessment of convergent validity, the PAS was correlated, using Pearson's *r*, with a parallel therapist scale (TAS). Coefficients ranged from 0.14 (inadequate) to 0.94 (adequate), the median being 0.76 (adequate) (probabilities not reported)	
In an examination of the scale's factor structure, a correlational measure of association and an average linkage algorithm was calculated, from which three clusters emerged. Each cluster consisted of items that, theoretically, might be expected to go together[1]	
Face	The PAS was devised from previously published measures (e.g. Bales, 1950; Goodman and Dooley, 1976; see ref 1) and revised on the basis of suggestions from the authors' colleagues
	Items were written to avoid theoretical jargon and/or the need for complex clinical inferences in favour of specific operationally defined actions
Content	As Face validity, above
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	Between the PAS and TAS, Pearson's *r* ranged from 0.14 to 0.94, median = 0.76 (probabilities not reported)
Discriminant	No details
Factor structure	Three clusters emerged from a cluster analysis:
	'reactions to therapist plus expressive-avoidance' includes items 1, 8, 10, 11, 12 and 19
	'working through the stress event' includes items 3, 4, 5, 6, 9, 15, 17, 23 and 24
	'termination' includes items 22, 25, 26 and 27
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	24
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1981, *Psychotherapy: Theory, Research and Practice*
Web or scanning options	No details
Training details	In the initial assessment of the scale, raters underwent approximately 12 hours of training before beginning the actual task

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Administration/process details	Clients' sessions were audio-recorded with their consent. Four sessions were rated from each therapy: an early, early–middle, late–middle and late session Instructions on the rating form ask the rater to indicate whether each of the listed patient actions occurred and, if so, to assess its emphasis in relation to the overall action in the session
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Likert-type. Each item has a response scale from 0 'did not do it' to 5 'major emphasis'
Normative data:	No details
Notes	
The PAS appears in Hoyt *et al*. (1981)[1]	
Résumé	
Strengths	The PAS has face and content validity, being devised from previously published measures and with suggestions from the authors' colleagues. The items were written to avoid theoretical jargon and/or the need for complex clinical inferences in favour of specific operationally defined actions[1]
Weaknesses	Validation work was conducted with only 25 clients, 21 of whom were women. The scale may be of limited value to a service provider as it is specifically for brief dynamic therapy While the median Finn's *r* are adequate for both inter-rater and test–retest reliability, the ranges are large (0.45 to 0.95 and 0.54 to 0.99, respectively), which includes inadequate and partial correlations. Convergent validity (with the parallel TAS) assessments using Pearson's *r* yielded similar results, i.e. the mean was adequate but the range was vast, from 0.14 to 0.96[1] Raters require approximately 12 hours of training before rating
Areas for further research	Further examination of psychometric properties. All assessments have been conducted with a small sample group, by the authors of the scale, and there are wide-ranging results
Primary reference	
1. Hoyt MF, Marmar CR, Horowitz MJ, Alvarez WF. The Therapist Action Scale and the Patient Action Scale: instruments for the assessment of activities during dynamic psychotherapy. *Psychother Theory Res Pract* 1981;**18**:109–16.	
Secondary references	
None	

P2
Penn Helping Alliance Questionnaire (HAq)

General details	
Author	Woody GE
Language	English
Country of publication/development	USA
Publication date	1983
Publisher	No details
Purpose and overview	
As a patient self-report measure, the HAq method assesses the extent to which the patient experiences the therapist and the therapy as helpful	
Theoretical orientation	Psychodynamic,[2] drug counselling,[4] supportive–expressive,[4] cognitive-behavioural,[4] and individual psychotherapy[7]
Population details	See notes below
Perspective	Patient self-report/observer-rated
Measure used by	Therapists, researchers
Other versions	Penn Helping Alliance Questionnaire – Revised Penn Helping Alliance Questionnaire – Dutch translation Penn Helping Alliance Rating Scale Penn Helping Alliance Counting Signs Method
Notes	*Patients:* Veteran: non-psychotic, methadone hydrochloride-maintained drug-dependent patients in the VA-Penn Project (Woody *et al*. 1983)[1,4] Outpatients: of the participants, only 163 were a new sample, rest were taken from a previous study. 83 male and 148 female participants, median age 27. Most patients had neurotic problems and mild character disorders. 76% single, 95% white. Most patients were in or had graduated from college[2] Depressed patients[3] Caucasian therapist-client dyads (mean age 30 years)[7] 121 adult patients (mean age 38 years) at admission and discharge from a psychiatric day-treatment unit[8] 48 methadone-maintained male opiate addicts with antisocial personality disorder[9] Borderline personality disorder patients (aged 17–35 years)[10] *Practitioners:* Therapist[1,2,7,10,11,13] paraprofessional,[2] counsellor,[1,9] psychotherapist,[1,4,9] drug therapist[4]
Areas of therapist–patient interaction addressed: Map	
Roles: expert/authority/leader Patient engagement: expectation/preferences Framework: collaborative/participative/involving Outcomes: general satisfaction; achieving a working relationship Inferred from full scale	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Dimensions	
Helping alliance	Eight items. Extent to which the patient experiences the therapist as providing, or able to provide, needed help (e.g. 'I believe that my therapist is helping me')
Collaboration	Three items. Extent to which the patient experiences therapy as a collaborative effort (e.g. 'I feel that I am working together with the therapist in a joint effort; we are on the same team')
Reliability	
There is no published evidence of the reliability of the HAq. A formal reliability study is underway for the HAq method[1]	
Split-half	No details
Internal consistency	No details
Inter-rater	No details
Test–retest	No details
Validity	
Content validity has been suggested for the HAq. The predictive validity of the HAq ranges from partial to adequate. The HAq demonstrates adequate convergent validity with the California Psychotherapy Alliance Scale (CALPAS), the Working Alliance Inventory (WAI) and a measure of therapists' personal qualities. Factor analysis supports the two dimensions of the HAq	
Face	No details
Content	The California Psychotherapy Alliance Scale – Patient Commitment (CALPAS-PC) and the HAq helping relationship scale come nearest to capturing the dimension of positive, forward-moving work identified as the confident collaboration factor in this study[2]
Criterion (a) concurrent	No details
Criterion (b) predictive	Support for this comes from examining a number of correlations derived from both during-treatment and pretreatment measures. The HAq method predicted 7-month outcomes in the VA-Penn study with impressive success at the high end of the range of correlations found in previous studies. The correlations, which were all statistically significant ($p < 0.01$), ranged from 0.51 in legal status, to 0.58 in psychological status, to 0.70 in employment status and 0.72 in drug use (Luborsky *et al*., 1985)[1,4]
Construct	No details
Convergent	HAq correlated with CALPAS $r = 0.74$ and with WAI $r = 0.74$ ($p < 0.0001$)[2] Tichenor and Hill (1989) used the Haq-1 as an observer rating and showed partial overlap with other rated measures; its correlation with the WAI was 0.71 ($p < 0.05$), but correlations were lower with the other scales: VTAS (0.51) and the CALPAS (0.34). However, the self-report versions of the three scales did not, and perhaps do not correlate significantly with the observer-rated versions[3] Therapist's personal qualities (interest in helping patients, psychological skill and health as rated by three independent judges) were highly correlated with the HAq (0.74)[4] HAq, purity and personality measures were moderately related to each other (mean $r = 0.63$)[4]
Discriminant	No details
Factor structure	The HAq has two factors by parallel analysis and the Scree test, with eigenvalues of 5.93 and 1.39, accounting for 67% of the variance. The two factors consist, respectively, of help items (items 1–5) and relationship items (items 6–10). Item 11 did not load on the analysis. The HAq items have the greatest tendency to load highly on several factors. Factored by itself, the HAq splits into two overall factors, helpfulness and relationship to the therapist. The distribution of items on these factors suggests that patients have a different sense of what is involved in helping alliance and collaboration than the HAq's authors[1]

continued

Responsiveness	
Discriminative (between individuals)	HAq scores for the drug counselling group within the antisocial patients were significantly lower than those of the psychotherapy groups (Gerstley *et al*., 1988)[4]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	11
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1986, Guilford Press
Web or scanning options	No details
Training details	It is essential that raters be clinically experienced to use the HA methods
Administration/process details	A patient's score equals the sum of the subscale ratings
Support from measure developers	See Appendix V for the HAq methods manual in ref. 1
FAQ facility	No details
Precision	
Scale type	Ordinal. Likert. Each item is rated on a six-point Likert scale, with a range from +3 ('Yes, I strongly feel that it is true') to –3 ('No, I strongly feel that it is not true') Rated on a five-point scale[4]
Normative data	Refer to ref. 1
Notes	
The HAr is time-consuming and expensive to use, requiring typescripts, audiotapes or videotapes, and so Luborsky developed the Penn Helping Alliance Questionnaire (HAq). The questionnaire method is simpler and much more economical to use than the observer-rated method. It also provides different, perhaps more direct assessments of the therapeutic alliance, since the patient, not independent observers, does the assessments "The HAQ seems to be a 'quick scan' instrument to give a quick and global impression of the patients' perception of the quality of the working alliance with the therapist" (De Weert-Van Oene *et al*., 1999). Easy to complete (De Weert-Van Oene *et al*., 1999) The results of this study point to the need for substantial revision of alliance measures. "We recommend that the items from the HAQ not be used in alliance research. Except for its unique emphasis on help received, the HAq's questions are too general to discriminate aspects of alliance effectively" (ref. 2, p. 1335)	
Résumé	
Strengths	Simple and economical to use. Demonstrates generally adequate validity and responsiveness
Weaknesses	There is no published evidence of the reliability of the HAq The HAq, although capturing the collaborative helpfulness of effective therapy, is too general and non-specific to distinguish important aspects of the alliance[2] Raters need to be clinically experienced
Areas for further research	Investigations of the reliability of the HAq

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Primary references

1. Alexander LB, Luborsky L. The Penn Helping Alliance Scales. In Greenberg LS, Pinsof WM, editor. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford; 1986. pp. 325–66.
2. Hatcher RL, Barends AW. Patients' view of the alliance in psychotherapy: exploratory factor analysis of three alliance measures. *J Consult Clin Psychol* 1996;**64**:1326–36.
3. Luborsky L. A pattern-setting therapeutic alliance study revisited. *Psychother Res* 2000;**10**:17–29.
4. Luborsky L, Crits-Christoph P, Mintz J, Auerbach A. *Who will benefit from psychotherapy: predicting therapeutic outcomes*. New York: Basic Books; 1988.
5. Luborsky L, Diguer L, Krause E, Levine J, Blum A, Ermold J. An expanded Helping Alliance Counting Signs manual for sessions and its association with the Helping Alliance Questionnaire (HAQ-1) – a comparison across alliance methods. Unpublished manuscript; 1999.
6. Luborsky L, Levine J, Johnson S, Diguer L, McLellan AT, Seligman DA. *The helping alliance questionnaire (HAQ-1): a research digest and comparison with the HAQ-II*. Unpublished manuscript; 1999.

Secondary references

7. Bachelor A, Salame R. Participants' perceptions of dimensions of the therapeutic alliance over the course of therapy. *J Psychother Pract Res* 2000;**9**:39–53.
8. Dazord A, Gerin P, Seulin C, Duclos A, Amar A. Day treatment evaluation: therapeutic outcome after a treatment in a psychiatric day-treatment center: another look at the 'outcome equivalence paradox'. *Psychother Res* 1997;**7**:57–70.
9. Gerstley L, McLellan AT, Alterman AI, Woody GE, Luborsky L, Prout M. Ability to form an alliance with the therapist: a possible marker of prognosis for patients with antisocial personality disorder. *Am J Psychiatry* 1989;**146**:508–12.
10. Gunderson JG, Najavits LM, Leonhard C, Sullivan CN, Sabo AN. Ontogeny of the therapeutic alliance in borderline patients. *Psychother Res* 1997;**7**:301–9.
11. Kaufman M. Effects of therapist self-monitoring on therapeutic alliance and subsequent therapeutic outcome. *Clin Supervis* 2000;**19**:41–59.
12. Luborsky L, McLellan AT, Woody GE, O'Brien GP, Auerbach A. Therapist success and its determinants. *Arch Gen Psychiatry* 1985;**42**:602–11.
13. Roten YD, Gillieron E, Despland J.-N, Stigler M. Functions of mutual smiling and alliance building in early therapeutic interaction. *Psychother Res* 2002;**12**:193–212.

P3
Penn Helping Alliance Questionnaire – Revised (HAq-II)

General details	
Author	Luborsky L
Language	English
Country of publication/development	USA
Publication date	1996
Publisher	NA
Purpose and overview	
The HAq was limited by the presence of items that were explicitly assessing early symptomatic improvement and by the fact that all items were worded positively. The HAq-II was devised to address these limitations	
Theoretical orientation	Cognitive, individual drug counselling, group drug counselling and supportive–expressive dynamic therapy[1]
Population details	See below
Perspective	Patient self-report and therapist self-report
Measure used by	Researchers, practitioners
Other versions	Penn Helping Alliance Questionnaire (HAq)
Notes	*Patients:* Outpatients with a DSM-III-R diagnosis of cocaine dependence. Average age was 33 ± 6.6 years, 60% female. 56% Caucasian. 61% employed. 24% married or living with partner. 75% primarily crack users *Practitioners:* Counsellor,[1] therapist[1] Supportive–expressive (SE) dynamic therapists and cognitive therapists (CT). Selected by their training units on the basis of a combination of background education and training, letters of reference and two audiotaped samples of their therapy/counselling work. The SE and CT therapists recruited to this study had performed an average of 9.9 and 10.6 years of postgraduate clinical work, respectively. Drug counsellors could not exceed certain levels of qualification[1]
Areas of therapist–patient interaction addressed: Map	
Roles: protector; expert/authority/leader Framework: collaborative/participative/involving; convergent Therapist engagement: empathy/sensitivity; genuineness; listening Patient engagement: motivation; attraction Threats to the relationship: critical Outcomes: general satisfaction; working relationship Inferred from items of patient self-report scale; no items provided for therapist version	
Dimensions	
No details	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
Patient version: adequate internal consistency and test–retest reliability	
Therapist version: adequate internal consistency and partial test–retest reliability	
Split-half	No details
Internal consistency	Cronbach's alpha. The HAQ-II patient scale had the following internal consistency values; at session 2, 0.90 ($n = 174$); session 5, 0.90 ($n = 171$); and session 24, 0.93 ($n = 83$). HAQ-II therapist scale had internal consistency 0.93 ($n = 193$) at session 2, session 5 was 0.90 ($n = 169$) and session 24 was 0.91 ($n = 0.88$)[1]
Inter-rater	No details
Test–retest	Test–retest reliability coefficients for both measures, over a three-session timespan from session 2 to session 5; HAQ-II patient version (0.78, $n = 168$, $p < 0.001$), HAQ-II therapist version (0.56, $n = 166$, $p < 0.001$). Stability-over-time correlations were 0.34 ($n = 74$, $p < 0.005$) for patient version and 0.55 ($n = 78$, $p < 0.001$) for the therapist version[1]

Validity	
Both versions of the HAq-II have adequate convergent validity with the California Psychotherapy Alliance Scale (CALPAS) at various points of the course of therapy. The discriminant validity of the HAq-II is questionable	
A two-factor structure is supported, with factor 1 (positive therapeutic alliance) accounting for 43.3% of the variance and factor 2 (negative therapeutic alliance) accounting for 10.6% of the variance	
Face	To address the limitations of the HAq, Luborsky *et al*. (1996) deleted the six items reflecting early improvement and added 14 items that appeared to tap more fully the various aspects of the alliance as described by Bourdin (1979) and Luborsky (1976). Five of the new items related to the collaborative effort of therapist and patient, five additional items addressed the patient's perception of the therapist, one of the other added items dealt directly with the patient's motivation. In contrast to the previous version, the HAq-II included five items that were worded negatively. One of the most important changes introduced in this new version of the HAq is the attempt to eliminate items that directly reflect symptomatic improvement
Content	No details
Criterion (a) concurrent	HAq-I and -II have not yet been administered concurrently [1]
Criterion (b) predictive	No details
Construct	No details
Convergent	Correlations between HAq-II and CALPAS total filled out by patients and therapists at sessions 2, 5 and 24: session 2 HAq-II (P, $n = 197$) 0.59, (T, $n = 200$) 0.79; session 5 (P, $n = 182$) 0.68, (T, $n = 174$) 0.79; session 24 (P, $n = 92$) 0.69, and (T, $n = 87$) 0.75. All $p < 0.001$. Five of the items from each scale are virtually the same, and several others are very close in meaning[1]
Discriminant	The higher the alliance (as measured by the HAq-II and the CALPAS), the lower the amount of drug use during the same week. The correlations were significant for session 5 (–0.18, $p < 0.05$), although not for session 2[1]
Factor structure	Conducted a principal components analysis with a varimax rotation. Using the Scree test and criterion of eigenvalues greater than 1, three factors were extracted. Because the third factor consisted of only two items and explained only 6% of the variance, this factor was not retained. Factor 1 (positive therapeutic alliance) explained 43.3% of the variance, factor 2 (negative therapeutic alliance) explained 10.6% of the variance. At session 2 the correlation between factors 1 and 2 was found to be $r = 0.48$ ($n = 200$, $p < 0.001$). At session 5, the correlation was $r = 0.60$ ($n = 182$, $p < 0.001$); at session 24, r was 0.64 ($n = 87$, $p < 0.001$)[1]

continued

Responsiveness	
Discriminative (between individuals)	No relation between alliance measured by HAq-II and sociodemographic variables was found. HAq-II was not associated with intake measures of psychological functioning, psychiatric severity, drug use or depression level[1]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	19
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1996, *Journal of Psychotherapy Practice and Research*
Web or scanning options	No details
Training details	No details
Administration/process details	No details
Support from measure developers	Patient version is reproduced in Appendix A of Luborsky *et al*. (1996)[1]
FAQ facility	No details
Precision	
Scale type	Ordinal. Likert. Each item is rated on a six-point Likert scale (1 = I strongly feel it is not true, 6 = I strongly feel it is true). Negatively worded items are reverse scored
Normative data	No details
Notes	
Luborsky (2000, p. 25): a combination of the two main types of procedures should be used for assessment of the alliance: a self-rating questionnaire method (e.g. the HAQ-II, the revised and expanded alliance measure) and a clinical observer rating method (e.g. the HAcs or HAr)	
Résumé	
Strengths	Patient version: adequate internal consistency and test–retest reliability. Therapist version: adequate internal consistency and partial test–retest reliability Both versions of the HAq-II have adequate convergent validity with the California Psychotherapy Alliance Scale (CALPAS) at various points of the course of therapy
Weaknesses	The discriminant validity of the HAq-II is questionable
Areas for further research	Further experience with the HAq-II in non-addicted patients would increase confidence in the generalisablity of the present findings[1]
Primary references	
1. Luborsky L, Barber JP, Siqueland L, Johnson S, Najavits LM, Frank, A, *et al*. The revised Helping Alliance Questionnaire (HAq-II): psychometric properties. *J Psychother Pract Res* 1996;**5**:260–71. 2. Luborsky L, Levine J, Johnson S, Diguer L, McLellan AT, Seligman DA. The helping alliance questionnaire (HAQ-1): a research digest and comparison with the HAQ-II. Unpublished manuscript; 1999.	
Secondary reference	
3. Luborsky L. A pattern-setting therapeutic alliance study revisited. *Psychother Res* 2000;**10**:17–29.	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

P4
Penn Helping Alliance Rating Scale (HAr)

General details	
Author	Luborsky L
Language	English
Country of publication/development	USA
Publication date	1986
Publisher	NA
Purpose and overview	
Intended to estimate the helping alliance. Focuses on two types of alliance (see Dimensions)	
Theoretical orientation	Humanistic,[2,9] cognitive-behavioural,[2] psychoanalytic,[2,6–8] supportive–expressive,[6] behavioural,[8] person-centred,[8] psychodynamic,[9] bioenergetic[9] and various/range[3,4]
Population details	See Notes
Perspective	Independent observer
Measure used by	Practitioners, researchers
Other versions	Helping Alliance Counting Signs Method (HAcs)
Notes	*Patients:* 37 women (mean age 31.2 years) and ten men (mean age 28.9 years). 64% single, 34.8% had some university education, 43% unemployed, 35% had psychoneuroses, 33% interpersonal problems and 28% personality disorders. Client's pretherapy overall psychological functioning on the Global Assessment Scale was 'moderate symptoms or generally functioning with some difficulty'. High symptomatology was also found on the Psychiatric Symptom Index[2] Participants' reported mean days of cocaine use in the 30 days before entering treatment was 12.98 (SD 8.83)[3] Data drawn from psychotherapy sessions that were part of a randomised clinical trial evaluating treatment for co-morbid cocaine and alcohol dependence[4] Ten of the most and ten of the least improved among the 73 in the Penn Psychotherapy Project. Improvement was based upon two moderately highly correlated (0.76) composite outcome measures: rated benefits and residual gain[6] All patients were non-psychotic. The ten least improved patients who were treated for at least 25 sessions were chosen from the 73 audiotaped cases in the Penn Psychotherapy Project on the basis of two correlated composite outcome measures. Mean age 26 years; 13 were female, all non-psychotic[7] Data for this study were taken from eight cases of brief (12–20 sessions) of psychotherapy[8] *Practitioners:* Graduates[2] Clinicians[3,4] Psychologist[3,4] Psychiatrist[3,6] Psychotherapist[6] Paraprofessional[6,7] Four male and four female therapists ranged in age from 34 to 78 years, with 5–42 years of postdoctoral experience[8]

continued

	Raters: Highly experienced psychoanalysts[7] Six doctoral students in counselling or clinical psychology served as raters for the observer-rated working alliance measures[8]
Areas of therapist–patient interaction addressed: Map	
Therapist engagement: empathy/sensitivity; warmth; respect Framework: collaborative/participative/involving; complementary General satisfaction: satisfaction Achieving a working relationship: working alliance (task, affective bond, goals) Changing view of self with others: narrative truths; modification of working models; corrective emotional experience Inferred from dimensions information provided below and from full scale	
Dimensions	
Type 1 alliance	Type I reflects psychoanalytic focus on the client's affective bond (warm, supportive, helpful) with the therapist (Freud, 1958). The patient's perception of the therapist as providing needed help.[8] The manual contains six subtypes of helping alliance (HA) type I: (a) the patient believes the therapist or therapy is helping; (b) the patient feels changes since the beginning of the treatment; (c) the patient feels a rapport with the therapist; (d) the patient feels optimism and confidence that the therapist and treatment can help; (e) the patient feels that the therapist is warm and supportive; and (f) the patient feels that the therapist respects and values him or her
Type 2 alliance	Type 2 appears more closely related to Bordin's concept of mutual agreement on tasks and goals (Bordin, 1979). The patient's experience of treatment as a collaboration with the therapist on the goals of treatment.[8] Four HA type 2 subtypes are included: (a) the patient experiences himself as working together with the therapies in joint effort; (b) the patient shares with the therapist similar conceptions about the sources of his problems; (c) the patient demonstrates qualities which are similar to those of the therapist, especially those connected with the tools for understanding; and (d) the patient expresses his or her belief that he or she is increasingly able to cooperate with the therapist in terms of understanding his or her own behaviour
Reliability	
The internal consistency of the HAr ranges from partial to adequate, with the majority of the results suggesting adequate reliability. There is similar pattern of findings for inter-rater reliability. In both cases, it is only the fixed/random effects ICC results that suggest partial reliability Test–retest reliability shows no significant changes and therefore is demonstrated to be adequate, but one study showed partial test-retest reliability	
Split-half	No details
Internal consistency	The Pearson's r correlation of the ten subscales, using a pair of raters, ranged from 0.75 to 0.88, with most in the 0.80's[1] The correlation between HA Type 1 and HA Type 2 was greater than 0.70[2] Coefficient alpha for HA was 0.96[7] Coefficient alpha of 0.93. Luborsky *et al*. (1983) reported a coefficient alpha of 0.96 for the total scale[8] Fixed effects ICC of 0.69. Tichenor and Hill (1989) reported that the Penn had high internal consistency (0.93)[3]
Inter-rater	See Internal consistency[1] 0.41 random-effects ICC[3] For the eight sessions that were rated by all raters, 0.71 was the random-effect ICC estimate[4]
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	Tang and DeRubeis[10] found inter-rater correlation of 0.60. The average score composite reliability of the two raters (Allen and Yen, 1979) was estimated as 0.75. In a sample of 19 depressed patients and using an enlarged manual to increase the range of the measure (Luborsky *et al.*, 1999), the agreement of the two judges with each other was a kappa of 0.73[5]
	Correlations were in the 0.8 to 0.9 range (Mintz *et al.*, 1979)[6]
	For the ten scales in the HA type 1 + HA type 2, correlations ranged from 0.75 to 0.88, with most in the 0.80s[7]
	0.71 using the formula for consistency between raters (Shrout and Fleiss, 1979)[8]
Test–retest	HA scores of the two early sessions are moderately consistent with scores of the two late sessions for the HAr method, ($r = 0.53$, $p < 0.05$)[6]
	The early and late sessions' ratings had a similar level of helping alliance ratings. Essentially, there was no significant gain from early to late sessions in either HA1 or HA2 scores. Likewise, an analysis of variance using session scores showed no significant early vs late effects[7]
Validity	
The predictive validity of the HAr has been demonstrated in numerous studies, and where figures are given, they range from partial to adequate validity	
The construct validity has been demonstrated through correlating HA type 1 and HA type 2 and HAr with HAcs, suggesting adequate validity	
Partially adequate to adequate convergent validity has been found for the HAr through testing against a number of measures (see below)	
The HAr has demonstrated adequate discriminant validity with the Holmes and Rahe Life Change Scale but not with conceptually different dimensions of the Therapeutic Alliance Rating Scale (TARS) and Vanderbilt Psychotherapy Process Scale (VPPS) measures	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	For HA type 1 scores the test for treatment (more vs less improved) was significant ($F = 5.9$, df = 1, 1, 18, $p < 0.05$). The test for treatment stage (early vs late) and the test for treatment outcome-by-treatment stage interaction were not significant. Analyses of HA type 2 scores showed similar results[1]
	From the vantage point of clients' alliance perceptions, therapist-provided HA type 1 proved to be the strongest predictor of improvement, accounting for 49% and 48% of client-rated positive change in multidimensional psychological functioning and target complaints, respectively. When therapists' ratings of the alliance were used, HA type 2, this was a relatively consistent contributor to predicting outcome[2]
	Correlation between alliance and outcome for the Penn is 0.50 ($p < 0.001$) for all treatments, 0.63 ($p < 0.001$) for CBT, and 0.42 ($p < 0.001$) for TSF[4]
	Significant predictive power was found for HAr. The combination of three simple outcome rating scales by the therapist – success, satisfaction and improvement (SSI) – was significantly predicted by early positive HAr (0.49, $p < 0.05$)[6]
	Predictions of outcomes by helping alliance ratings were significant. For HA1, the test for treatment outcomes (more vs less improved) was significant ($F = 5.9$, df = 1, 18, $p < 0.05$). The test for treatment stage (early vs late) and the test for the treatment outcome-by-treatment stage interaction were not significant (using actual outcome scores did not improve the discrimination). The analyses of HA2 scores showed similar results. The relationship of helping alliance ratings with outcomes is also expressed correlationally. The Health–Sickness Rating Scale (HSRS) ratings together with the helping alliance ratings provided impressive multiples[7]
	Luborsky *et al.* (1983) reported that both HA1 and HA2 in the early sessions were related to outcome[8]

continued

Validity	
Construct	HA type 1 scores were highly correlated with HA type 2 scores (0.91). HAcs and HAr were significantly correlated for both early and later session ratings[1]
Convergent	HA type 1 correlated relatively highly with Vanderbilt Psychotherapy Process Scale (VPPS) variables of therapist exploration and therapist warmth and friendliness. HA type 2 correlated highly with VPPS therapist exploration. From the therapist's viewpoint, HA type 1 and VPPS therapist warmth and friendliness correlated highly. HA type 2 was strongly associated with the VPPS patient hostility. The correlation coefficients between the HAr, Therapeutic Alliance Rating Scale (TARS) and VPPS ranged from 0.71 to 0.86 ($p < 0.001$), accounting for between 50 and 74% of the shared variance. A multitrait multimethod procedure correlated the same dimensions across the three instruments; correlations ranged from 0.46 to 0.83. Theoretically convergent dimensions were significantly more highly associated with each other than either theoretically divergent dimensions, supporting the convergent and discriminant validities of the alliance dimensions covered[2]
	The Penn correlated with the following measures: California Psychotherapy Alliance Scale (CALPAS) 0.54 ($p < 0.001$), Vanderbilt Therapeutic Alliance Scale (VTAS) 0.47 ($p < 0.001$), Working Alliance Inventory – Observer Rated (WAI-O) 0.50 ($p < 0.001$), Client Rated (WAI-C) 0.32 (ns) and Independent Rater (WAI-I) 0.38 (ns)[3]
	The Penn intercorrelated with VTAS 0.49 ($p < 0.001$), WAI-O 0.53 ($p < 0.001$), WAI-C 0.36 (ns) and WAI-T 0.44 ($p < 0.001$)[4]
	The session-based HAr measure and the session-based HAcs measure did agree, even though each had been scored by two different pairs of independent judges. The high level of agreement implies some validity for the method (Luborsky *et al*., 1983, 1988, Table 12–3). They agreed more highly for the late sessions (0.83, $p < 0.001$) than for the early sessions (0.57, $p < 0.01$), perhaps because in the late sessions, the outcome of the treatment might have been more evident in what the patient and the therapist said in the session[5]
	In a sample of 19 depressed patients, the correlation of the HAcs with the HAr was 0.51 (significant at the 0.05 level, two-tailed)[5]
	The HAcs and HAr significantly correlated with each other for both the early and late sessions' ratings. They agreed more highly for the late sessions (0.83, $p < 0.001$) than for the early sessions (0.57, $p < 0.01$). The fact that scores on these two methods scored by two different pairs of judges showed moderate agreement may imply some validity for the methods. Early HAr correlated 0.85 ($p < 0.001$) with early Therapist Facilitating Behaviors by the Rating Method (TFBr); 0.76 ($p < 0.001$) with late TFBr[6]
	Intercorrelations between the measures in the study yielded the following results (all significant at $p < 0.05$ level or better): HA1 and HA2 0.91, HA1 and Therapist Facilitative Behaviors scale (TFB 1 0.61, and TFB2 0.85, HA1 and resistance –0.60 and with insight 0.67. HA2 correlated with TFB1 0.50, TFB2 0.74, resistance –0.69, and insight 0.83[7]
	The Penn correlated with: CALPAS 0.34 (ns), VTAS 0.51 (ns), WAI-O 0.71 ($p < 0.05$), WAI-C 0.02 (ns), and WAI-T 0.20 (ns)[8]
Discriminant	Conceptually different dimensions of the HAr, TARS and VPPS measures were correlated; correlations ranged from 0.01 to 0.80 (absolute scores), with an average correlation of 0.39[2]
	An 'Amount of life change' measure based upon the Holmes and Rahe Life Change Scale (Holmes and Rahe, 1967) correlated –0.52 ($p < 0.05$) with HAr[6]
Factor structure	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	The early and late sessions' ratings had a similar level of helping alliance ratings. Essentially, there was no significant gain from early to late sessions in either HA1 or HA2 scores, although the HA2 scores for the more improved patients showed a non-significant increase, which had been anticipated. Likewise, an analysis of variance using session scores showed no significant early vs late effects. The correlation of early vs late sessions for HA1 was 0.57 ($p < 0.01$). The correlations were high for the more improved patients (0.69), but insignificant for the less improved patients
Acceptability	
Number of items	Ten
Administration method	Questionnaire and interview
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1986, Guilford Press
Web or scanning options	No details
Training details	It is essential that raters be clinically experienced to use the HA methods. For the normative study (Alexander *et al*., 1986), two independent judges were given a session of training in the use of the manual by a collaborator
Administration/process details	In the HAr method, the scoring unit and contextual unit are the same: the first 20-minute segment of a therapy session. The choice of the 20-minute segments is based on theoretical and empirical evidence. The beginning rather than the middle or the end of the session was chosen because it was felt that the judge needed to know all that had happened in the session thus far to assess adequately evidence of the patient's experience of a helping alliance. The 20-minute session was chosen as a compromise between the whole session, which, although desirable, would have been too costly and time-consuming to transcribe, and the more typical 5-minute unit in psychotherapy process research, which would have been too short for judging a relationship variable such as the helping alliance (Mintz and Luborsky, 1971)[1]
Support from measure developers	See Appendix II for the HAr Method Manual in Alexander *et al*. (1986)[1] or in Morgan *et al*. (1982)[7]
FAQ facility	No details
Precision	
Scale type	Ordinal. Each item is rated on a ten-point Likert-type scale reflecting the degree to which each item was present
Normative data	The normative sample to which the HAr has been applied was identical sessions from 20 patients, the ten most and the 120 least improved among the 73 patients in the Penn Psychotherapy Project (Luborsky *et al.,* 1980; Morgan *et al.*, 1982; Luborsky *et al.*, 1983). 18 therapists treated these 20 patients in supportive–expressive (SE) psychoanalytically orientated psychotherapy, recently described in a manual (Luborsky, 1984). All patients were non-psychotic, most of whom came for treatment at the outpatient clinic of the Hospital of the University of Pennsylvania

continued

Notes

Although the Helping Alliance Rating Method (HAr) seems simpler, the evidence so far is that the Helping Alliance Counting Method (HAcs) is not difficult to use and the psychometric properties of the HAcs appear to be somewhat better than those of the HAr.[5] The HAcs usually attained slightly higher predictive correlations than the HAr[6]

Since the HAr method is less time-consuming than the HAcs method, it has a practical advantage

Unlike the HAcs, the HAr method allows judges more freedom to use their clinical acumen by permitting them to examine the entire segment to be scored and rating it as a unit. The HAr method represents the conversion of the HAcs subscales into ten-point, Likert-type scales

The HAr is time-consuming and expensive to use, requiring typescripts, audiotapes or videotapes, and so Luborsky developed the Penn Helping Alliance Questionnaire (HAq). The questionnaire method is simpler and much more economical to use than the observer-rated method. It also provides different, perhaps more direct assessments of the therapeutic alliance, since the patient, not independent observers, does the assessments

Résumé

Strengths	The internal consistency of the HAr ranges from partial to adequate, with the majority of the result, suggesting adequate reliability. There is a similar pattern of findings for inter-rater reliability
	Test–retest reliability shows no significant changes and therefore is demonstrated to be adequate, but one study showed partial test–retest reliability
	The HArs has demonstrated some adequate convergent, discriminant, construct and predictive validity data
	Less time consuming than the HAcs method
	Allows judges more freedom to use their clinical acumen
Weaknesses	The psychometric properties of the HAcs appear to be somewhat better than those of the HAr[5]
	Time-consuming and expensive to use
Areas for further research	

Primary references

1. Alexander LB, Luborsky L. The Penn Helping Alliance Scales. In Grenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook*. 1986. pp. 325–66.
2. Bachelor A. Comparison and relationship to outcome of diverse dimensions of the helping alliance as seen by client and therapist. *Psychotherapy* 1991;**28**:534–49.
3. Cecero JJ, Fenton LR, Frankforter TL, Nich C, Carroll KM. Focus on therapeutic alliance: the psychometric properties of six measures across three treatments. *Psychotherapy* 2001;**38**:1–11.
4. Fenton LR, Cecero JJ, Nich C, Frankforter TL, Carroll KM. Perspective is everything: the predictive validity working alliance instruments. *J Psychother Pract Res* 2001;**10**:262–8.
5. Luborsky L. A pattern-setting therapeutic alliance study revisited. *Psychother Res* 2000;**10**:17–29.
6. Luborsky L, Crits-Christoph P, Alexander LB, Margolis, M, Cohen M. Two helping alliance methods for predicting outcome of psychotherapy. *J Nerv Mental Dis* 1983;**171**:480–91.
7. Morgan RW, Luborsky L, Crits-Christoph P, Curtis H, Solomon J. Predicting the outcomes of psychotherapy by the Penn Helping Alliance Rating Method. *Arch Gen Psychiatry* 1982;**39**:397–402.
8. Tichenor V, Hill CE. A comparison of six measures of working alliance. *Psychother* 1989;**26**:195–9.

Secondary references

9. Bachelor A, Salame R. Participants' perceptions of dimensions of the therapeutic alliance over the course of therapy. *J Psychother Pract Res* 2000;**9**:39–53.
10. Tang TZ, DeRubeis RJ. Sudden gains and critical sessions in CBT for depression. *J Consult Clin Psychol* 1999;**67**:894–904.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

P5
Psychotherapy Process Inventory (PPI)

General details	
Author	Baer P
Language	English
Country of publication/development	USA
Publication date	1980
Publisher	NA

Purpose and overview	
The PPI measures psychotherapy process over an extended period (macroanalysis), with an average rating given across all sessions with a particular patient	
Theoretical orientation	Pan-theoretical
Population details	Adults with diagnosis of personality disorders and neuroses[1]
Perspective	Therapist
Measure used by	Therapists/research therapists
Other versions	None
Notes	

Areas of therapist–patient interaction addressed: Map

Therapy context: boundaries; influence

Roles: expert/authority/leader

Individual differences: defensive style/repression

Patient engagement: motivation; commitment; Intentions

Framework: collaborative/participative/involving; structuring; directive

Therapeutic techniques: feedback

Threats to the relationship: defensive; hostility/anger; resistance; confrontations

Outcomes: general satisfaction

Information derived from example items

Dimensions	
Therapeutic participation	e.g. The patient tried to change his/her behaviour
Resistance	e.g. The patient was competitive with the therapist
Directive support	e.g. The therapy involved giving the patient 'homework'
Dysphoric concerns	e.g. The patient tended to be self-derogatory

Reliability	
The PPI demonstrated adequate internal consistency across all four dimensions	
Split-half	No details
Internal consistency	Coefficient alphas for the four dimensions were as follows: Therapeutic participation: 0.92 Resistance: 0.87 Directive support 0.83 Dysphoric concerns: 0.79

continued

Inter-rater	ICC among six raters (interns) observing eight sessions of a single patient was 0.43
Test–retest	No details
Validity	
The PPI demonstrates predictive validity with regard to therapeutic outcome, with the factors of therapeutic participation and dysphoric concerns being the strongest predictors	
Factor analysis supports a four-factor structure which accounts for 40.5% of the variance	
Face	The PPI was derived from representative sources in the clinical literature and from discussions with practising psychotherapists who represented a variety of theoretical orientations. Criteria for construction and choice of items included (1) capability of eliciting differences among therapists; (2) minimising the degree of inference required for rating; (3) coverage of broad range of theoretical positions and concepts framed in non-technical language; (4) focus on concepts common to a variety of theoretical positions
Content	Low item correlations showed that no item appeared to duplicate another item
Criterion (a) concurrent	No details
Criterion (b) predictive	Treatment outcome (using a seven-point scale – not specified) was significantly related ($p < 0.05$) to ratings of psychotherapeutic process (as measured by PPI): • The better the estimate of treatment outcome the higher the score on the factor of therapeutic participation • Patients with the best outcomes had significantly higher scores on the factor of dysphoric concerns
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	The 74 items were factor analysed using orthogonal (varimax) rotation, with the requirement of a potential factor accounting for 5% total variance being used in extracting factors. Four factors were extracted accounting for 40.5% of the total variance. The factor loadings on each factor exceeded 0.50. The factors were: therapeutic participation (concentrating on patient motivation), resistance (concentrating on the relationship between the therapist and patient), directive support (concentrating on the therapist) and dysphoric concerns (concentrating on the content of patients' verbalisations)
Responsiveness	
Discriminative (between individuals)	Factor scores for the five therapists who rated nine patients or more were compared. There were significant variations among therapists on all factor scores ($p < 0.05$) The PPI distinguished among patients who had good and poor outcomes (see Predictive validity section) Patients classified as normal according to the MMPI (Minnesota Multiphasic Personality Inventory) had significantly higher scores on therapeutic participation and patients who were classified as hysteric had significantly higher resistance scores than those who were classified as character disordered
Evaluative (within individual across time)	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Acceptability	
Number of items	74
Administration method	Therapist-completed questionnaire
Time taken to complete	30 minutes per PPI
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1980, *Psychological Reports*
Web or scanning options	No details
Training details	No details
Administration/process details	The instruction to the therapists specified that each rating was to be an 'average' over the number of sessions for which the patient had been seen
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Each of the 74 items was rated on a five-point ordinal scale of either frequency or intensity and some items were rated for both
Normative data	No details
Notes	
The PPI has been used to examine the process of child psychotherapy[2]	
Résumé	
Strengths	Adequate internal consistency. Factors discriminate among therapists and clients in a clinically meaningful way
Weaknesses	Inadequate inter-rater reliability
Areas for further research	Further research on concurrent/convergent validity with other established process measures/inventories
	Further work on cross-validation with more diverse samples

Primary reference

1. Baer PE, Dunbar PW, Hamilton JE, Beutler LE. Therapists' perceptions of the psychotherapeutic process: development of a psychotherapy process inventory. *Psychol Rep* 1980;**46**:563–70.

Secondary references

2. Gorin SS. The prediction of child psychotherapy outcome: factors specific to treatment. *Psychother Theory Res Pract Train* 1993;**30**:152–8.
3. Kolb DL, Beutler LE, Davis CS, Crago M, Shanfield SB. Patient and therapy process variables relating to dropout and change in psychotherapy. *Psychother Theory Res Pract Train* 1985;**22**:702–10.

P6
Psychotherapy Process Q-Set (PPQS)

General details	
Author	Jones EE
Language	English
Country of publication/development	USA
Publication date	2000
Publisher	NA
Purpose and overview	
The Q-Set comprises three types of items: (1) those describing patient and attitude and behaviour; (2) those reflecting the therapist's actions and attitudes; and (3) those attempting to capture the nature of the interaction in the dyad or the climate or atmosphere of the encounter	
Each item is printed on separate cards to permit easy arrangement and rearrangement. The items are sorted into nine piles ranging from 'least characteristic' (category 1) to 'most characteristic' (category 9), with a middle pile (category 5) used for items deemed either 'neutral' or 'irrelevant'. The number of cards sorted into each pile must conform to a normal distribution (ranging from 5 at the extremes to 18 in the middle or 'neutral' category. The items provide a standard format of clinically meaningful units that observers can use to clarify and describe the process material under study	
Theoretical orientation	Developed for study of psychoanalysis, but has been used for study of cognitive behavioural and interpersonal therapies[1,2]
Population details	Depressed adults with diagnosis of depression,[1,2] undergraduates with assertiveness problems in close relationships,[3] professional women with stress response syndrome and experience of traumatic life event/loss[4]
Perspective	Clinical judges
Measure used by	Psychoanalytic clinicians,[6] psychotherapists,[1,2] graduates,[3] social workers, psychologists and psychiatrists[4]
Other versions	No details
Notes	No details
Areas of therapist–patient interaction addressed: Map	
Therapy context: boundaries; influence; power/coercion	
Roles: expert/authority/leader	
Individual differences: attachment styles; defensive style/repression	
Therapist engagement: empathy sensitivity; support/tolerance; listening; hope/encouragement; praise/affirmation	
Patient engagement: motivation; expectation/preferences; attraction; commitment; intentions	
Framework: collaborative/participative/involving; controlling; structuring; directive; challenging; focused	
Therapeutic techniques: transference; responsiveness; exploration; ruptures/repair; feedback	
Non-verbal communication: laughter/humour; paralinguistics	
Threats to the relationship: defensive; critical; hostility/anger; fear; sexual involvement; resistance; confrontations	
Outcomes: emotional expression; changing view of self with others	
Information inferred from items	
Dimensions	
None	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
Across studies the PPQS has shown adequate internal consistency and inter-rater reliability	
Split-half	No details
Internal consistency	Alpha coefficient: 0.82,[1] average item reliability: 0.82[12]
	Coefficient alpha reliabilities were 0.95 for cognitive behaviour therapy (CBT) and 0.96 for interpersonal psychotherapy (IPP)[2]
Inter-rater	Inter-rater reliabilities across five raters: 0.87[3]
	Inter-rater reliabilities for two judges ranged from 0.71 to 0.89[4]
	Mean inter-rater reliability 0.86; (range 0.68 to 0.90)[5]
	Mean inter-rater reliability 0.86, (range 0.58 to 0.95)
Test–retest	No details
Validity	
The PPQS was developed to ensure acceptable face and content validity. The instrument has mixed/limited evidence on predictive validity, but has been shown to be responsive to changes over time and to differentiate between different types of therapies	
Face	100 items that comprise the Q-Set represent an empirically guided selection from a pool of several hundred items gathered from existing process measures, as well as new items constructed by a panel of experts[6]
Content	Several versions of the Q-Set were tested in a series of pilot studies conducted on scores of video and audio-tapes of psychotherapy and psychoanalytic treatment hours. Items were eliminated if they showed little variation over subjects and therapy hours, were redundant or had low inter-rater reliability[6]
Criterion (a) concurrent	No details
Criterion (b) predictive	A hierarchical multiple regression analysis was conducted for each Q item. 27 Q items were significant predictors of therapy outcome. However, almost all significant findings were interaction effects, with Q items' value predictive of outcome in interaction with patient pretreatment disturbance level. Only one item ('patient achieves a new understanding or insight') predicted outcome independent of the seriousness of pathology (F of R^2 change $= 4.66$, $p < 0.05$)[4]
	Correlations with outcome measures ranged from 0.36 to 0.53 in CBT condition and 0.11 to 0.48 in IPP condition[2]
Construct	No details
Convergent	Two authors identified all possible PQS items that seemed similar to any of the Hill Counselor Verbal Response Category System (CVRCS) clusters. The resultant internal consistencies (alpha) of the proposed clusters were only included if alpha was >0.70. PQS items that corresponded to all seven CVRCS clusters were identified. None of the Q-Set clusters significantly correlated with the corresponding CVRCS cluster[3]
Discriminant	No details
Factor structure	Factorial validity for the Q-Set is irrelevant because the measure was constructed in a manner that insured independence among items. A factor analysis of the Q-Set, including various rotational possibilities, revealed an absence of factor structure which is desirable from the standpoint of Q methodology[4]
	Principal components factor analysis yielded distinct theoretical orientation factors with eigenvalues above 1.0 after varimax rotation, which together explained 70.9% of the variation in the correlations among the expert therapists[2]
	The Q-Set was constructed to minimise the emergence of general factors. In a factor analytic study based on two data sets which included 70 treatments, 130 separate treatment hours and 380 Q sorts, no clear factor structure was found[5]

continued

Responsiveness	
Discriminative (between individuals)	The instrument can differentiate between types of therapy: rational emotive, gestalt and client-centred. The ten items designated most and least characteristic for each form of therapy were presented to a group of five therapists familiar with those treatment modalities. The therapists successfully matched ($p < 0.001$) the sets of the Q items with the therapy from which they had been derived[5]
	48 of the 100 Q items significantly differentiated ($p < 0.05$) between IPP and CBT[2]
Evaluative (within individual across time)	The process of early sessions was remarkably similar to the process of sessions late in treatment, as evidenced by the fact that only four of the 100 Q items significantly differentiated between the early and late sessions ($p < 0.01$)[2]
Acceptability	
Number of items	100
Administration method	Judge-completed rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	2000, Enrico E. Jones
Web or scanning options	No details
Training details	Coding manual provided.[13] Minimal clinical experience (1 year of supervised therapy) allows judges to become reliable raters after a short period of training with the instrument
Administration/process details	After studying the transcript of a therapy hour clinical judges order the 100 items (printed separately on cards). The items are sorted into nine piles ranging on a continuum from least characteristic (category 1) to most characteristic (category 9). The middle pile (category 5) is used for items deemed neutral/irrelevant
Support from measure developers	Coding manual (see above)
FAQ facility	No details
Precision	
Scale type	Sorting procedure and rating scale: 100 statements sorted into nine categories. A predetermined distribution of the items among the nine categories. See ref. 13 for details
Normative data	No details
Résumé	
Strengths	Particularly suited for individual case study process research
	Has adequate reliability and validity
Weaknesses	Lengthy scoring and sorting procedure (100 items)
	Q-method cannot provide complete information on analytic discourse
	Q sorts impose a particular distribution to the items – can be constraining
Areas for further research	Further work to develop the PPIs responsiveness to change over time
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Primary references

1. Ablon JS, Jones EE. Psychotherapy process in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consul Clin Psychol* 1999;**67**:64–75.
2. Ablon JS, Jones EE. Validity of controlled clinical trials of psychotherapy: findings from the NIMH Treatment of Depression Collaborative Research Program. *Am J Psychiatry* 2002;**159**:775–83.
3. Heaton KJ, Hill CE, Edwards LA. Comparing molecular and molar methods of judging therapist techniques. *Psychother Res* 1995;**5**:141–53.
4. Jones EE, Cumming JD, Horowitz MJ. Another look at the nonspecific hypothesis of therapeutic effectiveness. *J Consult Clin Psychol* 1988;**56**:48–55.
5. Jones EE, Cumming JD, Pulos SM. Tracing clinical themes across phases of treatment by a Q-set. In Miller NE, Luborsky L, Barber JP, Docherty J, editors. *Psychodynamic treatment research: a handbook for clinical practice*. New York: Basic Books; 1993. pp. 14–36.
6. Jones EE, Windholz M. The psychoanalytic case study: toward a method for systematic inquiry. *J Am Psychoanal Assoc* 1990;**38**:985–1015.

Secondary references

7. Ablon JS, Jones EE. How expert clinicians' prototypes of an ideal treatment correlate with outcome in psychodynamic and cognitive-behavior therapy. *Psychother Res* 1998;**8**:71–83.
8. Grace M, Kivilighan DM, Kunce J. The effect of nonverbal skills training on counsellor trainee nonverbal sensitivity and responsiveness and on session impact and working alliance ratings. *J Counsel Dev* 1995;**73**:547–52.
9. Jones EE, Pulos SM. Comparing the process in psychodynamic and cognitive-behavioral therapies. *J Consult Clin Psychol* 1993;**61**:306–16.
10. Jones EE, Hall SA, Parke LA. The process of change: the Berkeley Psychotherapy Research Group. In Beutler L, Crago M, editors. *Psychotherapy Research: an International Review of Programmatic Studies*. Washington, DC: American Psychological Association; 1991. pp. 98–106.
11. Jones EE, Krupnick JH, Kerig PK. Some gender effects in brief psychotherapy. *Psychotherapy* 1987;**24**:336–52.
12. Jones EE, Parke LA, Pulos SM. How therapy is conducted in the private consulting room: a multidimensional description of brief psychodynamic treatments. *Psychother Res* 1992;**2**:16–30.
13. Jones EE. T*herapeutic action: a guide to psychoanalytic therapy*. Northvale: Jason Aronson; 2000.
14. Lambert MJ, Hill CE. Assessing psychotherapy outcomes and processes. In Bergin AE, Garfield SL, editors. *Handbook of psychotherapy and behavior change*. New York: Wiley; 1994. pp. 72–113.
15. Price PB, Jones EE. Examining the alliance using the Psychotherapy Process Q-Set. *Psychotherapy* 1998;**35**:392–404.

R1
Reasons for Ending Treatment Questionnaire (RETQ)

General details	
Authors	Garcia JA, Weisz JR
Language	English
Country of publication/development	USA
Publication date	2002
Publisher	NA
Purpose and overview	
The measure covers reasons why youths end treatment, and produces a score that indicates the likelihood of dropout from treatment (the higher the score, the greater the likelihood of dropout)	
Theoretical orientation	Child/adolescent therapy
Population details	Children/adolescents
Perspective	Therapist/parent
Measure used by	Child/adolescent psychology practitioners
Other versions	No details
Notes	Families were recruited from ten community clinics in California at the time of initial child intake assessment. RETQ reports were obtained after treatment had ended. The sample included 344 client families. All participants were included regardless of when their contact with clinic had ended. Client age range was 7–18 years (mean 11.73; SD 2.60); 63% were boys; 51% were Caucasian
Areas of therapist–patient interaction addressed: Map	
Therapy context: responsibilities	
Roles: advocate; protector	
Individual differences: level of functioning; problem complexity; social support	
Therapist engagement: empathy/sensitivity; warmth; listening	
Patient engagement: motivation; expectation/preferences; attraction; commitment; intentions	
Threats to the relationship: defensive; critical; hostility; resistance; confrontations; withdrawal	
Outcomes: satisfaction	
Information derived from items	
Dimensions	
Therapeutic relationship problems	15 items. Targeted at the therapist or the wider therapeutic team, e.g. The therapist didn't seem to understand
Family and clinical practical problems	Ten items, e.g. Someone in the family got sick, or appointments last too long
Staff and appointment problems	Seven items, e.g. The appointment interfering with the child's schooling
Time and effort concerns	Four items, e.g. We did not have enough time
Treatment not needed	Three items, e.g. I didn't really feel that my child had a problem
Money issues	Two items, e.g. The services cost too much

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
As estimated by Cronbach's alpha, the internal consistencies of all dimensions except for 'treatment not needed' met criteria for adequacy	
As estimated by Pearson correlations, the test–retest reliabilities of all dimensions met criteria for adequacy	
Split-half	NA
Internal consistency	Therapeutic relationship problems: 0.91 Family and clinic practical problems: 0.79 Staff and appointment problems: 0.75 Time and effort concerns: 0.71 Treatment not needed: 0.67 Money issues: 0.72
Inter-rater	No details
Test–retest	Therapeutic relationship problems: 0.91 Family and clinic practical problems: 0.84 Staff and appointment problems: 0.88 Time and effort concerns: 0.93 Treatment not needed: 0.76 Money issues: 0.93
Validity	
The predictive validity of the RETQ was partial: two of the six dimensions predicted dropout from treatment	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	Two of the dimensions (therapeutic relationship problems and money issues) were shown to differentiate significantly between those who would complete and drop out of treatment ($p < 0.05$)
Construct	The hypothesis that the scores of dropouts and completers would differ was demonstrated by two of the dimensions (see above)
Convergent	No details
Discriminant	No details
Factor structure	Unweighted least squares extraction was used, yielding six factors. The factors accounted for the following amounts of variance: Therapeutic relationship problems: 15.72% Family and clinic practical problems: 7.22% Staff and appointment problems: 5.96% Time and effort concerns: 4.74% Treatment not needed: 4.38% Money issues: 3.38%
Responsiveness	
Discriminative (between individuals)	Discriminated between dropouts and completers on two of the six dimensions
Evaluative (within individual across time)	NA
Acceptability	
Number of items	41
Administration method	Interview and questionnaire
Time taken to complete	No details
Flesch reading age	NA
Translations	No details
Access by ethnic minorities	No details

continued

Feasibility	
Copyright	2002, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	Therapists interviewed the parents after treatment had ended and then completed the RETQ
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Each item was rated on a Likert scale. Higher scores indicate greater dissatisfaction. No details of how the scale is scored
Normative data	No details
Notes	
The RETQ is derived from a longer questionnaire by Gould (1985)	
Résumé	
Strengths	The measure makes a valuable contribution to the study of attrition in practice settings: it has been shown to discriminate between completers and dropouts
Weaknesses	The measure has good internal consistency and test–retest reliability
	The interview and questionnaire format (41 items) is somewhat lengthy. This might cause difficulties in implementing the measure in some practice settings
Areas for further research	Further research on the measure's psychometric properties and application across more clinical settings
Primary reference	
1. Garcia JA, Weisz JR. When youth mental health care stops: therapeutic relationship problems and other reasons for ending youth outpatient treatment. *J Consult Clin Psychol* 2002;**70**:439–43.	
Secondary reference	
2. Gould MS, Shaffer D, Kaplan D. The characteristics of dropouts from a child psychiatry clinic. *J Am Acad Child Psychiatry* 1985;**24**:316–28.	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

S1
Session Evaluation Questionnaire (SEQ)

General details

Author	Stiles WB
Language	English
Country of publication/development	USA
Publication date	1980
Publisher	NA

Purpose and overview

The Session Evaluation Questionnaire (SEQ) measures individual counselling sessions along two evaluative dimensions (session depth/value and session comfort/ease), and a dimension of post session mood (positivity)

The SEQ is a revision of a measure used to assess the impact of self-analytic small group sessions[2,7] and is referred to as SEQ or SEQ Form 2 in the research literature

Theoretical orientation	Psychodynamic
Population details	Clinical adults in individual psychotherapy
Perspective	Client, therapist or independent rater
Measure used by	Researchers
Other versions	SEQ Form 3, SEQ Form 4
Notes	In developing the SEQ, the initial analyses were conducted with therapist and client ratings of 113 individual psychotherapy sessions[5]

Areas of therapist–patient interaction addressed: Map

Outcomes: satisfaction; safety/secure base

Therapist–patient interaction information derived from the SEQ items

Dimensions

Depth/value	Represented in the first half of the questionnaire ('This session was'), depth/value refers to the perceived power and value of the session
Smoothness/ease	Represented in the first half of the questionnaire ('This session was'), smooth/ease refers to the perceived comfort and pleasantness of the session
Positivity	Represented in the second half of the questionnaire ('Right now I feel'), positive feelings is a global dimension referring to postsession feelings of confidence, clarity, happiness and the absence of fear and anger

Reliability

Each of the SEQ's three indexes has demonstrated adequate internal consistency

Split-half	No details
Internal consistency	Internal consistency coefficients of the depth and smoothness indices for therapists, clients and independent raters ranged from 0.80 to 0.90[3]
Inter-rater	No details
Test–retest	No details

continued

Validity

The SEQ is a revision of a measure used to assess the impact of self-analytic small group sessions,[2,7] giving it face validity

Client, external rater and therapist ratings on the SEQ depth and smoothness dimensions have been assessed for predictive validity by correlations with four client-reported improvement measures taken from the beginning of therapy to 3-month follow-up. The improvement measures are the Symptom Checklist (SCL-90), Beck Depression Inventory (BDI), Self-Esteem Scale (SES) and a composite well-being measure (see ref. 1). Clients' ratings demonstrated inadequate predictive validity on all measures, as did external raters' depth ratings; external raters' smoothness ratings demonstrated partial predictive validity with three of the four improvement measures; and therapists' ratings demonstrated either partial or adequate validity with all but the SCL-90[1]

Factor analyses with clients' and therapists' ratings yielded very similar results. Two distinct factors emerged from the first half of the SEQ, and one factor emerged from the second half of the questionnaire[1]

Face	The SEQ items were developed from an earlier measure used to assess the impact of self-analytic small group sessions[2,7]
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	In assessments of predictive validity, client, external rater and therapist SEQ ratings of depth/value and smoothness/ease were correlated with client self-report improvement measures (from intake to 3-month follow-up; SCL-90, BDI, SES; and a composite well-being index), with the following results: Client ratings: no significant correlations with any of the four improvement measures[3] External rater ratings: depth ratings did not significantly correlate with any of the four improvement measures; smoothness correlations were not significant with SES; 0.41 ($p < 0.01$) with SCL-90, 0.42 ($p < 0.01$) with BDI and 0.39 ($p < 0.02$) with well-being[3] Therapist ratings: no significant correlations with SCL-90; depth correlated 0.32 ($p < 0.05$) with BDI, 0.43 ($p < 0.01$) with SES and 0.39 ($p < 0.02$) with well-being. Smoothness correlated 0.46 ($p < 0.01$) with BDI and SES, and 0.46 ($p < 0.01$) with well-being[3]
Construct	No details
Convergent	Correlations between client, therapist and independent rater perspectives on the depth and smoothness indices ranged from 0.06 to 0.45[3]
Discriminant	No details
Factor structure	*Factor analyses* Four factor analyses were conducted (one for each half of the SEQ for clients' and therapists' data), from which the following factor structures emerged: *First factor analyses of clients' and therapists' ratings* Two very similar factors emerged from both ratings. Factor 1 distinguished sessions that were valuable from those that were not and accounted for 33.2% (clients) and 39.2% (therapists) of the common variance. Seven items loaded onto the factor in each analysis with loadings between 0.51 and 0.81 (clients) and 0.65 and 0.83 (therapists)[2] Factor 2 distinguished smooth, pleasant sessions from unpleasant ones and accounted for 27.8% (clients) and 23.3% (therapists) of the total common variance. Four items loaded onto the factor in each analysis with loadings between 0.73 and 0.92 (clients) and 0.71 and 0.84 (therapists)[2] *Second factor analyses of clients' and therapists' ratings* One large factor emerged from both analyses, accounting for 54.3% (clients) and 53.8% (therapists) of the common variance. In each analysis all 11 items loaded between 0.60 and 0.85, with the positive and negative adjectives at opposite poles[2]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	Interscale correlations The correlations (coefficient alpha) between depth/value and smoothness/ease indexes were rs = –0.04 (clients) and –0.08 (therapists)[2] The global dimension of positive feelings was correlated (coefficient alpha) with the two session dimensions for clients and therapists individually with the following results (*n* range from 109 to 113): *For clients:* With depth/value: 0.43 ($p < 0.0001$) With smoothness/ease: 0.60 ($p < 0.0001$) With depth and smoothness combined: 0.74 ($R\ p < 0.0001$) *For therapists:* With depth/value: 0.70 ($p < 0.0001$) With smoothness/ease: 0.48 ($p < 0.0001$) With depth and smoothness combined: 0.82 ($R\ p < 0.0001$)[2]
Responsiveness	
Discriminative (between individuals)	One-way ANOVAs performed on the client and therapist scales (44 in total) found significant differences among therapists, Phillai's trace $V = 7.41$, approximate $F_{616,882} = 1.61$, $p < 0.0001$[2] Univariate ANOVAs of client and therapist ratings across three outcome groups (dropout, poor outcome and good outcome) found that: • Clients' depth ratings differentiated between the three outcome groups ($p < 0.05$, with good outcome clients reporting higher depth); smoothness ratings did not discriminate[1] Therapists' smoothness ratings differentiated between the three outcome groups ($p < 0.05$, with therapists reporting greater smoothness sessions with good outcome clients); depth ratings did not discriminate[1]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	22
Administration method	Rating scale
Time taken to complete	2 minutes
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1980, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	The SEQ is completed after the therapy session. Written directions are 'Please place an X on each line to show how you feel about this session'. The stem 'This session was' precedes the first 11 adjective pairs and 'Right now I feel' precedes the second set of 11 adjective pairs
Support from measure developers	No details
FAQ facility	No details

continued

Precision	
Scale type	Likert-type. Each item is a bipolar adjective pair, e.g. special – ordinary, deep–shallow, rated on a seven-point semantic differential scale (Osgood *et al*.; see ref. 2)
Normative data	No details

Notes

The SEQ items are reported in ref. 2

Other research uses of the SEQ include:

A study of the relationship between perceived session helpfulness and session evaluation in 11 clients (aged 23–42 years) and 11 counsellors (aged 24–47 years) during eight sessions of short-term counselling[4]

A study into the effect of non-verbal skills training on counsellor trainee non-verbal sensitivity and responsiveness and on session impact and working alliance ratings[5]

An examination of a method and framework within which psychotherapy providers can better match clients to therapists and techniques and enhance client satisfaction and treatment outcomes[6]

A study of client preference for styles of therapy in which 25 psychiatric day hospital clients rated videos of psychodynamic, cognitive-behavioural, humanistic, external and naive styles of therapy[8]

Résumé	
Strengths	The SEQ takes only 2 minutes to complete. Each of the SEQ's three indexes have demonstrated adequate internal consistency when scored by therapists, clients and independent raters (coefficients range from 0.80 to 0.90)[3] Parts of the SEQ have demonstrated responsiveness to different therapists and outcome groups. One-way ANOVAs performed on the client and therapist scales (44 in total) found significant differences among therapists, Phillai's trace $V = 7.41$, approximate $F_{616,882} = 1.61$, $p < 0.0001$.[2] Clients' depth ratings differentiated between three outcome groups (dropout, poor outcome and good outcome, $p < 0.05$). and therapists' smoothness ratings differentiated between the three outcome groups ($p < 0.05$)
Weaknesses	The SEQ has so far failed to establish predictive validity. When the SEQ depth and smoothness dimensions were assessed for their ability to predict outcome as measured by the Symptom Checklist (SCL-90), Beck Depression Inventory (BDI), Self-Esteem Scale (SES) or a composite well-being measure (see ref. 1), client ratings showed no significant relationships to outcome. Some external observer and therapist ratings also showed no relationships and the significant correlations were partial at best[1]
Areas for further research	The SEQ has since been revised (to SEQ 3 and then again to SEQ 4)

Primary references

1. Samstag LW, Batchelder ST, Muran JC, Safran JD, Winston A. Early identification of treatment failures in short-term psychotherapy: an assessment of therapeutic alliance and interpersonal behavior. *J Psychother Pract Res* 1998;**7**:126–43.
2. Stiles WB. Measurement of the impact of psychotherapy sessions. *J Consult Clini Psychol* 1980;**48**:176–85.
3. Stiles WB, Shapiro DA, Firth-Cozens JA. Correlations of session evaluations with treatment outcome. *Br J Clin Psychol* 1990;**29**:13–21.

Secondary references

4. Cummings AL, Barak A, Hallberg ET. Session helpfulness and session evaluation in short-term counselling. *Counsel Psychol Q* 1995;**8**:325–32.
5. Grace M, Kivlighan DM, Knunce J. The effect of nonverbal skills training on counsellor trainee nonverbal sensitivity and responsiveness and on session impact and working alliance ratings. *J Counsel Dev* 1995;**73**:547–52.
6. Herman SM. Therapist–client similarity on the multimodal structural profile inventory as a predictor of early session impact. *J Psychother Pract Res* 1997;**6**:139–44.
7. Stiles WB, Tupler LA, Carpenter JC. Participants' perceptions of self-analytic group sessions. *Small Group Behav* 1982;**13**:237–54.
8. Wanigaratne S, Barker C. Clients' preferences for styles of therapy. *Br J Clin Psychol* 1995;**34**:215–22.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

S2
Session Evaluation Questionnaire (SEQ) – Form 3

General details

Authors	Stiles WB, Snow JS
Language	English
Country of publication/development	USA
Publication date	1984
Publisher	NA

Purpose and overview

The Session Evaluation Questionnaire (SEQ) Form 3 is a revision of the Session Evaluation Questionnaire (SEQ) and measures the impact of individual counselling sessions along two evaluative dimensions (depth and smoothness) and two dimensions of postsession mood (positivity and arousal)

Form 3 differs from the SEQ with the addition of arousal, in accord with Russell's (1978, 1979; see refs 1 and 2) model of a ubiquitous two-dimensional affective space; and with the addition and substitution of a few scales to measure depth, smoothness and positivity more clearly

Theoretical orientation	Psychodynamic
Population details	Adult students and community residents with psychoneuroses and study/adjustment problems; adolescents
Perspective	Client and therapist
Measure used by	Researchers
Other versions	SEQ, SEQ Form 4
Notes	The sample employed in the development of the questionnaire consisted of 942 rated counselling sessions. Clients were adult students and community residents with psychoneuroses and study/adjustment problems[2]

Areas of therapist–patient interaction addressed: Map

Therapist–patient interaction information derived from the SEQ Form 3 items:

Framework
Outcomes: satisfaction; safety/secure base

Dimensions

Depth	Represented in the first half of the questionnaire ('This session was'), depth refers to the perceived power and value of the session
Smoothness	Represented in the first half of the questionnaire ('This session was'), smoothness refers to the perceived comfort and pleasantness of the session
Positivity	Represented in the second half of the questionnaire ('Right now I feel'), positivity is a global dimension referring to postsession feelings of confidence, clarity, happiness and the absence of fear and anger
Arousal	Represented in the second half of the questionnaire ('Right now I feel'), arousal refers to feelings of activation vs sleep

continued

Reliability	
Each of the four dimensions, from both client and therapist perspectives, has demonstrated adequate internal consistency	
Split-half	No details
Internal consistency	Coefficient alphas for depth, smoothness, positivity and arousal respectively were: 0.91 (counsellors) and 0.87 (clients) 0.89 (counsellors) and 0.93 (clients) 0.86 (counsellors) and 0.89 (clients) 0.82 (counsellors) and 0.78 (clients)[2]
Inter-rater	No details
Test–retest	No details
Validity	
Face and content validity Revisions to an earlier version of the SEQ to improve clarity and represent Russell's (1978, 1979; see refs 1 and 2) model of a ubiquitous two-dimensional affective space support the face and content validity of the SEQ Form 3	
Convergent validity A number of interscale correlations have been conducted at different levels of analysis (counsellor-level, client-level residuals and session-level residuals) from client and counsellor perspectives. The dimensions converge to various degrees depending upon the perspective and level of analysis[2]	
Client and counsellor perspectives on each dimension have been correlated at three levels of analysis. At counsellor level the perspectives did not significantly converge on any dimension; for client-level residuals, the perspectives significantly converged on smoothness only; and for session-level residuals, the perspectives significantly converged on all four perspectives[2]	
Discriminant validity Positivity and arousal dimensions are based on Russell's (1979, see ref. 2) model, which holds that the two dimensions are independent. However, when correlated, a moderate degree of convergence was found at certain levels of analysis, which indicates that the questionnaire failed this test of discriminant validity	
Factor structure Depth and smoothness correlated for session-level residuals, but not at the counsellor-level or for client level residuals[2]	
Face	See Content validity
Content	The arousal dimension was added to in accordance with Russell's (1978, 1979; see refs 1 and 2) argument of two ubiquitous bipolar mood dimensions. The questionnaire has also added and substituted a few scales from the SEQ to measure depth, smoothness and positivity more clearly[1]
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	For client and counsellor perspectives, each session dimension (depth and smoothness) was independently correlated with each postsession mood dimension (positivity and arousal) at three levels of analysis (counsellor-level, client-level residuals and session-level residuals). Coefficients ranged from not significant to 0.72 ($p < 0.001$)[2] Client and counsellor perspectives of each dimension were correlated at three levels of analysis (counsellor-level, client-level residuals and session-level residuals) with the following results: None of the correlations of counsellor-level means were significant[2] For client-level residuals the perspectives on smoothness correlated 0.27 ($p < 0.05$); depth, positivity and arousal correlations were not significant[2] For session-level residuals the perspectives on depth, smoothness, positivity and arousal correlated 0.20 ($p < 0.001$), 0.39 ($p < 0.001$), 0.21 ($p < 0.001$) and 0.10 ($p < 0.01$), respectively[2]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	Across three levels of analysis (counsellor-level, client-level residuals and session-level residuals) correlations between positivity and arousal ranged from not significant to 0.47 ($p < 0.001$). The convergence between the two dimensions does not correspond to the model on which they were based (Russell, 1979; see ref. 2) where the two affective states are independent[2]
Discriminant	Correlations were calculated between depth and smoothness at three levels of analysis, for both client and counsellor perspectives, with the following results: Depth–smoothness correlations of counsellor-level means were not significant for either client or counsellor ratings[2] Depth–smoothness correlations of client-level residuals were not significant for client or counsellor ratings[2] Depth–smoothness correlations of session-level residuals: 0.16 ($p < 0.001$) (client), 0.09 ($p < 0.05$) (counsellor)[2]
Factor structure	Eight separate principal factor analyses were conducted: one for each half of the questionnaire (regarding the session and possession mood); for client and therapist perspectives; at the session level ($n = 907$ to 919) and client level ($n = 74$) of analysis Each factor analysis yielded very similar results. For each analysis two factors emerged (depth and smoothness from analyses of the first half of the SEQ Form 3; positivity and arousal from analyses of the second half of the SEQ Form 3). Each pair of factors accounted for 63–77% of the variance[1] An exploratory factor analysis ($n = 17$) yielded similar results[1]
Responsiveness	
Discriminative (between individuals)	*t*-Tests for correlated samples showed that clients rated sessions as deeper ($t_{896} = 9.76$, $p < 0.001$), and smoother ($t_{894} = 4.26$, $p < 0.001$) than did therapists[2]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	24
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1984, British Psychological Society
Web or scanning options	No details
Training details	No details
Administration/process details	The SEQ is completed after the therapy session. Written directions are 'Please place an X on each line to show how you feel about this session'. The stem 'This session was' precedes the first 12 adjective pairs and 'Right now I feel' precedes the second set of 12 adjective pairs
Support from measure developers	No details
FAQ facility	No details

continued

Precision	
Scale type	Likert-type. Each item is a bipolar adjective pair, e.g. special–ordinary, deep–shallow, rated on a seven-point semantic differential scale (Osgood *et al*., see ref. 2)
Normative data	No details
Notes	
The SEQ Form 3 items are listed in ref. 2	
Other research uses of the SEQ Form 3 include:	
Node-link mapping in chemical-dependent adolescents[3]	
A multivariate analyses study of the relationship between node-link mapping in 169 methadone treatment clients and client and counsellor SEQ ratings[4]	
Validity assessment of the Session Impacts Scale (SIS)[5]	
An exploration of the history of the working alliance over time[6]	
Résumé	
Strengths	942 counselling sessions were used in the development of the questionnaire.[2] The questionnaire is short with 24 items. All four dimensions, from both client and therapist perspectives, have demonstrated adequate internal consistency (coefficient alphas range from 0.78 to 0.93)[3]
Weaknesses	At certain levels of analysis, there is a degree of convergence between arousal and positivity. While it is moderate (up to 0.47, $p < 0.01$) it is contrary to the theory on which the scales are based[2]
Areas for further research	Further assessment of psychometric properties, including independent work. The primary research on the SEQ so far has been conducted by its authors

Primary references

1. Stiles WB, Snow JS. Dimensions of psychotherapy session impact across sessions and across clients. *Br J Clin Psychol* 1984;**23**:59–63.
2. Stiles WB, Snow JS. Counseling session impact as viewed by novice counselors and their clients. *J Counsel Psychol* 1984;**31**:3–12.

Secondary references

3. Collier CR, Czuchry M, Dansereau DF, Pitre U. The use of node-link mapping in the chemical dependency treatment of adolescents. *J Drug Educ* 2001;**31**:305–17.
4. Dansereau DF, Dees SM, Greener JM, Simpson DD. Node-link mapping and the evaluation of drug abuse counseling sessions. *Psychol Addict Behav* 1995;**9**:195–203.
5. Elliott R, Wexler MM. Measuring the impact of sessions in process experiential therapy of depression: the Session Impacts Scale. *J Counsel Psychol* 1994;**41**:166–74.
6. Horvath AO, Marx RW. The development and decay of the working alliance during time-limited counselling. *Can J Counsel* 1990;**24**:240–60.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

S3
Session Evaluation Questionnaire (SEQ) – Form 4

General details

Authors	Stiles WB, Reynolds S, Hardy GE, Rees A, Barkham M, Shapiro DA
Language	English
Country of publication/development	England
Publication date	1994
Publisher	NA

Purpose and overview

The Session Evaluation Questionnaire (SEQ, Version 4) measures individual psychotherapy sessions in terms of session evaluation, postsession mood and therapist evaluation

The questionnaire has been developed from the SEQ and SEQ Form 3. The SEQ Version 4 differs with the addition of the therapist evaluation items and that its dimensionality was assessed with a British (as opposed to American) sample[2]

Theoretical orientation	Not specified
Population details	Clinical adults in individual therapy
Perspective	Client and therapist
Measure used by	Researchers
Other versions	SEQ, SEQ Form 3
Notes	Initial assessment of the SEQ Version 4 was conducted with a sample of 2414 client-rated British psychotherapy sessions[2]

Areas of therapist–patient interaction addressed: Map

Therapist–patient interaction information derived from the SEQ Version 4 items:

Roles: good object

Therapist engagement: warmth

Framework

Outcomes: satisfaction; safety/secure base

Dimensions

Depth/value	Represented in the first half of the questionnaire ('This session was'), depth/value refers to the perceived power and value of the session
Smoothness/ease	Represented in the first half of the questionnaire ('This session was'), smooth/ease refers to the perceived comfort and pleasantness of the session
Positivity	Represented in the second half of the questionnaire ('Right now I feel'), positive feelings is a global dimension referring to postsession feelings of confidence, clarity, happiness and the absence of fear and anger
Arousal	Represented in the second half of the questionnaire ('Right now I feel'), arousal refers to feelings of activation vs sleep
Good therapist	The dimension was added to the SEQ to assess the client's evaluation of the therapist. Three bipolar adjectives items (skilful–unskilful, cold–warm and trustworthy–untrustworthy) are preceded by the stem 'Today I felt my therapist was'

continued

Reliability	
All four dimensions of the SEQ Version 4 have demonstrated adequate internal consistency	
Split-half	No details
Internal consistency	Coefficient alpha for depth, smoothness, positivity, arousal and good therapist were 0.90, 0.92, 0.90, 0.80 and 0.77, respectively[2]
Inter-rater	No details
Test–retest	No details
Validity	
The SEQ Version 4 has face and content validity as it was developed from the SEQ Form 3 The SEQ Version 4 depth and arousal dimensions demonstrated partial predictive validity when correlated with client-rated working alliance and treatment outcome, respectively[1] Principal components extraction with varimax rotation at session level and client level revealed distinct depth and smoothness factors for the session evaluation items, positivity and arousal factors for the postsession mood items, and a good therapist factor for the therapist evaluation items[2]	
Face	See Content
Content	The SEQ Version 4 was developed from the SEQ Form 3[2]
Criterion (a) concurrent	No details
Criterion (b) Predictive	The correlation between depth (initial session) and client-rated working alliance (fourth session) was 0.34 (*p* not reported)[1] Higher arousal ratings in the middle sessions of brief therapy partially predicted better treatment outcome (0.41, *p* not reported)[1]
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	Factor structure was assessed with factor analyses and interscale correlations Six principal components extractions with varimax rotation were conducted with client ratings: three two each at session level ($n = 2360$ to 2397) and client levels ($n = 210$) for the session evaluation items, the postsession mood items and therapist evaluation items *Session evaluation items at session level* Factor 1 smoothness had five of the 12 items with loadings from 0.80 to 0.85 and a final eigenvalue of 4.29[2] Factor 2 depth had five of the 12 items with loadings from 0.73 to 0.82 and a final eigenvalue of 3.46[2] *Session evaluation items at client level* Factor 1 smoothness had five of the 12 items with loadings from 0.87 to 0.90 and a final eigenvalue of 4.89[2] Factor 2 depth had five of the 12 items with loadings from 0.85 to 0.91 and a final eigenvalue of 4.66[2] *Postsession mood items at session level* Factor 1 positivity had five of the 12 items with loadings from 0.77 to 0.84 and a final eigenvalue of 3.85[2] Factor 2 arousal had five of the 12 items with loadings from 0.56 to 0.79 and a final eigenvalue of 2.98[2] *Postsession mood items at client level* Factor 1 positivity had five of the 12 items with loadings from 0.84 to 0.93 and a final eigenvalue of 4.97[2] Factor 2 arousal had five of the 12 items with loadings from 0.71 to 0.87 and a final eigenvalue of 3.41[2]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	Therapist evaluation items One factor emerged with all three items loading onto it from both the session and client levels of analysis, called good therapist. At session and client levels, respectively, item loadings were from 0.69 to 0.80 and 0.77 to 0.94 and final eigenvalues were 1.71 and 2.37[2] *Interscale correlations were conducted at session and client levels* Depth and smoothness correlation coefficients were 0.06 ($p < 0.05$) (session) and 0.28 $p < 0.001$(client)[2] Positivity and arousal correlation coefficients were 0.09, (ns), (session) and 0.10 ($p < 0.001$) (client)[2] Good therapist and arousal correlation coefficients were 0.08 ($p < 0.001$) (session) and not significant (client)[2] Good therapist correlated with the other three dimensions at both levels of analysis between 0.25 and 0.62 ($p < 0.001$)[2]
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	27
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1994, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	For each item scale respondents are instructed to 'please circle the appropriate number to show how you feel about this session'
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Likert-type. Seven-point bipolar adjective scales. Higher scores indicate greater depth, smoothness, positivity and arousal. The stems for the session evaluation, postsession mood and therapist evaluation items, respectively, are: 'This session was,' 'Right now I feel' and 'Today I feel my therapist was'
Normative data	No details
Notes	
The SEQ Version 4 items are listed in ref. 2 A comment paper on the SEQ Version 4 discusses the issues of retest reliability, the connection of session-level measures to research strategies based on other units of analyses, and suggested uses of the questionnaire[1]	

continued

Research uses of the SEQ Version 4 include:

An examination of the interaction between therapeutic alliance and in-session process during the assessment phase of treatment using a collaborative therapeutic assessment model proposed by Finn and Tonsager (1997)[3]

An investigation into the relationship between client recall of sessions and effectiveness ratings of sessions[4]

An investigation of session evaluation and type of participant-recalled important event in novice counsellor dyads and experienced counsellor dyads during nine sessions of short-term counseling[5]

An examination of intellectual empathy and empathic emotion in relation to therapist pre-session mood and clients' session evaluations[6]

An examination of the relationship between client pre-session level of distress and client rating of the effectiveness of individual counselling sessions[7]

An investigation of the across-session patterns of session impact in the treatments of 117 depressed clients who were randomly allocated to eight or 16 sessions of cognitive-behavioural or psychodynamic-interpersonal therapy[8]

Investigations of the relationships between the therapeutic alliance (client and therapist rated), perceived curative factors and evaluations of therapy sessions[9]

An exploration of how volunteer clients evaluated therapist competence and how these evaluations are related to session outcome, treatment outcome and client satisfaction[10]

A study that investigated the relation of client–counsellor evaluation of initial interview to client return for another session[11]

Résumé

Strengths	A sample of 2414 clients were employed in the initial assessment of the questionnaire.[2] All of the five dimensions have adequate internal consistency (coefficient alphas range from 0.77 to 0.90)[2]
Weaknesses	Arousal did not adequately predict outcome and depth did not adequately predict client-rated working alliance[1]
Areas for further research	Further assessment of psychometric properties

Primary references

1. Mallinckrodt B. Session impact in counseling process research: comment on Elliott and Wexler (1994) and Stiles *et al*. (1994). *J Counsel Psychol* 1994;**41**:186–90.
2. Stiles WB, Reynolds S, Hardy GE, Rees A, Barkham M, Shapiro DA. Evaluation and description of psychotherapy sessions by clients using the session evaluation questionnaire and the session impacts scale. *J Counsel Psychol* 1994;**41**:175–85.

Secondary references

3. Ackerman SJ, Hilsenroth MJ, Baity MR, Blagys MD. Interaction of therapeutic process and alliance during psychological assessment. *J Person Assess* 2000;**75**:82–109.
4. Cummings AL, Hallberg ET, Slemon A, Martin J. Participants' memories for therapeutic events and ratings of session effectiveness. *J Cogn Psychother* 1992;**6**:113–24.
5. Cummings AL, Slemon AG, Hallberg ET. Session evaluation and recall of important events as a function of counselor experience. *J Counsel Psychol* 1993;**40**,156–65.
6. Duan C, Kivlighan DM Jr. Relationships among therapist presession mood, therapist empathy, and session evaluation. *Psychother Res* 2002;**12**:23–37.
7. Jones WP, Markos PA. Client rating of counselor effectiveness: a call for caution. *J Appl Rehabil Counsel* 1997;**28**:23–28.
8. Reynolds S, Stiles WB, Barkham M, Shapiro DA, Acceleration of changes in session impact during contrasting time-limited psychotherapies. *J Consult Clin Psychol* 1996;**64**:577–86.
9. Svensson B, Hansson L. Relationships among patient and therapist ratings of therapeutic alliance and patient assessments of therapeutic process: a study of cognitive therapy with long-term mentally ill patients. *J Nerv Ment Dis* 1999;**187**:579–85.
10. Thompson BJ, Hill CE. Client perceptions of therapist competence. *Psychother Res* 1993;**3**:124–30.
11. Tryon GS. Session depth and smoothness in relation to the concept of engagement in counseling. *J Counsel Psychol* 1990;**37**:248–53.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

S4
Session Impacts Scale (SIS)

General details

Author	Elliott R
Language	English
Country of publication/development	USA
Publication date	1994
Publisher	NA

Purpose and overview

This measure assesses the impact, in the view of the client, of psychotherapy sessions. Unlike other impact scales (e.g. the Session Evaluation Questionnaire), the SIS measures the specific content rather than the general emotional quality of clients' reactions to sessions. It was thus developed from clients' open-ended descriptions of significant therapy events. The measure assesses impact in two major ways: as 'helpful impacts' and 'hindering impacts'. The 'helpful impacts' can be further divided into 'task impacts' and 'relationship impacts'. Therefore, the measure can be seen as having two or three dimensions of impact

Theoretical orientation	Psychodynamic–interpersonal; cognitive-behavioural; process-experiential
Population details	Adults: professional/managerial/white collar with differing degrees of depression;[2] major depressive or related disorders[1]
Perspective	Client
Measure used by	Psychotherapists/research therapists
Other versions	None
Notes	

Areas of therapist–patient interaction addressed: Map

Therapy context: boundaries; influence; power/coercion; responsibilities

Roles: confidant; protector; attachment figure

Individual differences: problem complexity

Therapist engagement: empathy/sensitivity; warmth; genuineness; respect; openness; listening; hope/encouragement; praise/affirmation

Patient engagement: motivation; commitment; intentions

Framework: reciprocal; collaborative/participative/involving; challenging

Therapeutic techniques: responsiveness/receptiveness/attunement; exploration; ruptures/repair; feedback

Threats to the relationship: intrusive; defensive; critical; hostility/anger; fear; resistance; withdrawal

Outcomes: achieving a working relationship; emotional expression; changing view of self with others

Information derived from items

Dimensions

Helpful impacts (task and relationship impacts)	Ten items
Hindering impacts	Six items

continued

Reliability	
The internal consistency of the SIS was generally adequate, with only the hindering impacts scale displaying partial reliability	
Split-half	No details
Internal consistency	Coefficient alpha values were as follows:[1] Hindering impacts: 0.67 (with a low loading item removed) Helpful impacts: (0.92) Task impacts: (0.84) Relationship impacts: (0.91) Coefficient alpha values varied from 0.78 to 0.90 for the five dimensions (although one dimension had only one item, and therefore did not produce an internal reliability score)[2]
Inter-rater	NA
Test–retest	No details
Validity	
The SIS demonstrates partial concurrent validity with the SEQ	
None of the dimensions correlated with the arousal dimension of the SEQ, therefore providing support for the measure's discriminant validity	
Factor analysis supports the two- and three-factor solution, but Stiles *et al*. have also found a five-factor solution	
Face	Starting from content and cluster analyses of client's open-ended descriptions of significant events within sessions, Elliott and colleagues devised a taxonomy of 16 impacts
Content	See above
Criterion (a) concurrent	The dimensions of the SIS were correlated with the depth, smoothness and positivity dimensions of the SEQ.[1] See study for full details of correlations • Significant correlations with the client scales of the SEQ ranged from 0.22 to 0.58 ($p < 0.01$) • None of the SIS scales displayed correlations above 0.30 with therapist SEQ[1] Two sets of scores were analysed at the 'session level', i.e. looking at the pattern of results for different characteristics of session (e.g. those characterised by more depth), and the 'client level', i.e. looking at patterns of results for different characteristics of client (e.g. those reporting more depth). The session-level results were unaffected by mean differences among clients, therapists or client–therapist pairings, by conducting analysis on session-level deviation scores. The client-level results were unaffected by mean differences among therapists, by conducting analysis on client-level deviation scores[2] The correlations reported in ref. 2 relate to the five factors of understanding, problem solving, relationship, unwanted thoughts and hindering impacts The vast majority of correlations between SIS dimensions and SEQ dimensions are significant, although these significant correlations go as low as 0.06 owing to the large sample size. SIS's positive impact indexes (understanding, problem solving and relationship) are correlated with SEQ depth index at client and session levels (0.44 to 0.72)[2]
Predictive	Hindering impacts predicts dropout from therapy (see Responsiveness)
Construct	The hypothesised two- and three-factor solution was supported. The hypothesis that the SIS dimensions would not correlate with the arousal scale of the SEQ was supported. See Discriminant validity[1]
Convergent	See Concurrent
Discriminant	The SIS dimensions showed no significant correlations with the arousal dimension for either the therapist- and client-rated SEQ[1]
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Factor structure	Exploratory principal axis factor analysis with varimax rotation replicated the SIS structure quite successfully, three factors with eigenvalues >1 accounting for 59% of variance. On rotation, 'the unwanted thoughts' item loaded lowly on 'hindering impacts factor' (0.3), and three items cross-loaded on 'task impacts' and 'relationship impacts' at >0.40. Apart from this, the three-factor structure was a good replication. Apart from an inadequate loading of the unwanted thoughts item, when a two-factor structure was sought (hindering impacts and helpful impacts, the latter incorporating task impacts and relationship impacts), the structure again was a good replication of the measure's structure. However, confirmatory factor analyses provided solutions which deviated significantly from the data[1] Principal components extraction was followed by varimax rotation. The 17th item of the scale 'other important aspects' was excluded from the factor analysis. Two or three factors emerge with eigenvalues exceeding 1, yet a five-factor solution was chosen: (1) understanding (2) problem solving, (3) relationship, (4) unwanted thoughts and (5) hindering impacts. Item loadings on these factors range from 0.61 to 0.77 for the session-level data, and 0.67 to 0.89 for the client-level data[2]
Responsiveness	
Discriminative (between individuals)	Dropouts report more hindering impacts than completers[1]
Evaluative (within individual across time)	A Pearson product-moment correlation was calculated between helpful impacts and session progression. A small to medium trend of increasing scores over sessions was found (3–4% variance), but was not significant[1]
Acceptability	
Number of items	17
Administration method	Self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1994, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	For each item clients are asked to rate the item on the basis of the descriptor that best fits their experience. Each item includes a label and a short paragraph description
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Each of the 17 items/descriptors is rated on a five-point scale ranging from (1) (not at all) to 5 (very much) in terms of how it fits with the client's experience of the session
Normative data	No details

continued

Notes	
A 17th item, 'other important aspects', is open ended, and may be excluded from the measure. It is important to re-emphasise that the measure can be interpreted as having either two or three dimensions: helpful impacts and hindering impacts, or splitting the helpful impact items into task impacts and relationship impacts	
Résumé	
Strengths	Adequate internal consistency. Promising evidence for concurrent/convergent validity with the SEQ. Factor analysis supports the SIS's construct validity. Discriminant validity demonstrated. SIS responsive
Weaknesses	Mixed evidence relating to the factor structure of the SIS
Areas for further research	Further research to clarify the factor structure of the SIS
	Research to clarify the measure's responsiveness to change over time

Primary references

1. Elliott R, Wexler MM. Measuring the impact of sessions in process–experiential therapy of depression: the session Impacts Scale. *J Counsel Psychol* 1994;**41**:166–74.
2. Stiles WB, Reynolds S, Hardy GE, Rees A, Barkham M, Shapiro DA. Evaluation and description of psychotherapy sessions by clients using the Session Evaluation Questionnaire and the Session Impacts Scale. *J Counsel Psychol* 1994;**41**:175–85.

Secondary reference

3. Hill CE, Kellems IS. Development and use of the Helping Skills Measure to assess client perceptions of the effects of training and of helping skills in sessions. *J Counsel Psychol* 2002;**49**:264–72.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

T1
Therapeutic Alliance Scales for Children

General details	
Authors	Shirk SR, Saiz CC
Language	English
Country of publication/development	USA
Publication date	1992
Publisher	NA
Purpose and overview	
The purpose of the Therapeutic Alliance Scales for Children is to assess the child's experience of the therapeutic relationship. The scales are designed to distinguish between the child's affective experience of therapy and their collaboration with therapeutic tasks (e.g. talking about problems, expressing feelings)	
Theoretical orientation	The scales follow Bordin's (1979, see ref. 1) concept of a multifaceted alliance and are relevant across theoretical orientations
Population details	Children in the clinical population
Perspective	Self-report and therapist rated
Measure used by	Researchers, psychologists and psychiatrists
Other versions	The scales have client and therapist versions
Notes	The participants in the pilot study were 62 children, aged 7–12. Participants evidenced serious psychopathology and were receiving inpatient treatment including several weekly sessions of individual therapy. The prevailing individual therapy orientation was psychodynamic[1]
Areas of therapist–patient interaction addressed: Map	
Framework: collaborative/participative/involving	
Outcomes: working alliance: affective bond; goals	
Derived from a general description of the measure	
Dimensions	
Bond	The child's positive orientation towards therapy
Negativity	The child's negative orientation towards therapy
Verbalisation	The child's collaboration on tasks, verbalisation of problems
Reliability	
In the pilot study, adequate reliability was measured with Cronbach's alpha. Adequate reliability was demonstrated by all three therapist perspective scales and the child perspective bond and negativity scales. There was partial support for the reliability of the child perspective verbalisation scale[1]	
Split-half	No details
Internal consistency	The alphas for the bond, negativity and verbalisation scales were 0.88, 0.72 and 0.87 (therapist perspective) and 0.72, 0.74 and 0.67 (child perspective) respectively (probabilities not given) Eight items were initially written in parallel for the child and therapist. One therapist item was dropped owing to a low item-total correlation
Inter-rater	No details
Test–retest	No details

continued

Validity	
Factor structure was demonstrated in the pilot study as the intercorrelations among the alliance subscales were all significant and in the expected direction (of varying strength). That is, verbalisation correlated positively with bond and negatively with negativity for both perspectives and bond and negativity were inversely related. Also, child and therapist perspectives for the affective quality of the relationship converged moderately, although not so much as to be interchangeable. This indicates that each participant makes a unique contribution to the understanding of the affective quality of the relationship[1]	
Convergent validity was indicated for therapist (but not child) ratings of bond and negativity, which were related to 'global' (a separate therapist rating of participation in therapy adapted from the Menninger Collaboration Scale, see ref. 1), indicating convergent validity[1]	
The more positive children felt towards therapy, the more likely they were to discuss their problems and feelings. However, as all the measures were taken at the same time, the causal direction is unclear[1]	
Face	No details
Content	The scales' items were solicited from experienced clinical psychologists and child psychiatrists[1]
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	Therapist global ratings of participation correlated with therapist bond and negativity 0.57 ($p < 0.001$) and –0.28 ($p < 0.01$) respectively
Discriminant	No details
Factor structure	In earlier research, an exploratory factor analysis of the original scale items yielded a three-factor solution. Two factors referred to orientation to therapy (positive and negative) and a third to therapeutic tasks[1]
	There was a moderate degree of convergence between the child and therapist perspectives for the affective quality of the therapeutic relationship, although the results indicate that the two perspectives are not interchangeable[1]
	Verbalisation correlated with bond and negativity 0.26 ($p < 0.05$) and –0.49 ($p < 0.001$) (child perspective) and 0.45 ($p < 0.001$) and –0.34 ($p < 0.01$) (therapist perspective), respectively[1]
	The bond and negativity correlation coefficients for the child and therapist perspectives were –0.57 ($p < 0.001$) and –0.50 ($p < 0.001$), respectively[1]
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	Eight items for the child perspective and seven for the therapist refer to affective orientation. There are no details on the number of items in the verbalisation scale
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details of time to complete or reading age, but the authors were guided by the need to make the scales relatively simple and brief
Translations	No details
Access by ethnic minorities	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Feasibility	
Copyright	1992, Cambridge University Press
Web or scanning options	No details
Training details	No details
Administration/process details	In the pilot study, therapists and clients completed the questionnaires during the child's third week of hospitalisation. A staff member who was not their therapist administered the scales to the children
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Each item is a statement, e.g. 'I like spending time with my doctor' is on the child perspective bond scale. There are no details as to whether the response is given nominally or on a Likert-type scale
Normative data	No details
Résumé	
Strengths	The scales were devised from Bordin (1979). Both scales are short and were designed to be simple. Of the three dimensions for both perspectives, all but the child verbalisation have adequate internal consistency. Therapist bond ratings adequately converged with a separate therapist rating of participation. Bond and negativity correlation coefficients are adequate and negative (as would be expected)[1]
Weaknesses	Verbalisation on the child form does not have adequate internal consistency.[1] Therapist-rated negativity did not adequately converge with a separate therapist rating of participation[1] Verbalisation did not adequately correlate with bond or negativity, from either perspective, where it was expected to[1]
Areas for further research	Further examination of psychometric properties, including independent work. The scales have been assessed against one other measure and the scales' authors so far have conducted all assessments

Primary reference

1. Shirk SR, Saiz CC. Clinical, empirical and developmental perspectives on the therapeutic relationship in child psychotherapy. *Dev Psychopathol* 1992;**4**:713–28.

Secondary references

None

T2
Therapeutic Bond Scales

General details	
Authors	Saunders SM, Howard KI, Orlinsky DE
Language	English
Country of publication/development	USA
Publication date	1989
Publisher	NA
Purpose and overview	
To assess the quality of the therapeutic relationship from the patient's perspective. The therapeutic bond is composed of three aspects: working alliance, empathic resonance and mutual affirmation. The scales were developed to measure these aspects and the therapeutic bond as a whole. The Therapeutic Bond Scales were extracted from the Therapy Session Report (TSR) questionnaire (Orlinsky and Howard, 1966, 1986b) which was designed as a general survey of the patients' intrasession experiences. The bond scales were developed on a conceptual basis and then subjected to psychometric revision to achieve maximum reliability	
Theoretical orientation	Psychodynamic
Population details	Clinical adults. 113 psychotherapy outpatients took part in the development study,[1] attending the Northwestern Memorial Hospital's Institute of Psychiatry. The typical patient was single, white female, aged 25–35 with some college education. Patients were self-referred and treated for a range of mild to moderate psychological disorders. Via screening interview, all patients were determined to be appropriate for psychodynamically orientated, intensive, individual therapy
Perspective	Patient self-report
Measure used by	Psychotherapists. In the development study there were psychologists, psychiatrists and social workers; the majority were in some stage of training but had considerable additional experience
Other versions	None
Notes	The median number of sessions received was 26[1]
Areas of therapist–patient interaction addressed: Map	
Therapist engagement	
Patient engagement	
Framework	
Outcomes: achieving a working relationship; emotional expression	
Information derived from items	
Dimensions	
Working alliance (WA)	The working alliance subscale addresses patient motivation for coming to the session and patient motivation for returning to the next session
Empathic resonance (ER)	The empathic resonance subscale refers to a quality of communication between patient and therapist that depends on their compatibility in the range and style of expressiveness and understanding
Mutual affirmation (MA)	The mutual affirmation subscale reflects care, respect and commitment to the other person's welfare that the patient and the therapist may evoke in and feel for one another
Global bond scale	The global bond scale is a composite of the three subscales
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
The development study[1] showed adequate reliability for the three subscales and partial reliability for the global bond scale	
Split-half	No details
Internal consistency	The global bond scale's internal reliability was 0.62. The working alliance scale had a reliability (alpha coefficient) of 0.72. The empathic resonance scale's reliability was 0.77. The mutual affirmation scale's reliability was 0.87
Inter-rater	No details
Test–retest	No details
Validity	
The development study[1] showed adequate predictive validity for the global bond scale in relation to patient ratings of session quality, and partial validity for the three subscales. Predictive validity as measured by ratings of termination outcome was less adequate	
Face	No details
Content	Items from the TSR were evaluated for appropriateness of inclusion on one of the Therapeutic Bond Scales, based on consensus among the three authors[1]
Criterion (a) concurrent	No details
Criterion (b) predictive	Two measures of therapeutic effectiveness were used: • Session quality was assessed using the patient's overall assessment of the session just completed (using the first item of the TSR). All the correlations between session quality and the bond scales were significant at $p < 0.001$ (WA: $r = 0.34$; ER: $r = 0.51$; MA: $r = 0.50$; Global $r = 0.60$) • Termination outcome scores were calculated from ratings of clinic files by independent judges, using the evaluation method developed by Tovian (1977). Termination outcome was correlated with the global bond score ($r = 0.19$, $p < 0.05$), but not the three bond subscales
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	Working alliance: 15 items; empathic resonance: 17 items; mutual affirmation: 18 items (composite global bond scale: 50 items)
Administration method	Questionnaire (written)
Time taken to complete	"The relative ease with which this instrument is administered, completed, and analyzed underscores its potential usefulness" (Saunders *et al*., 1989, p. 328). The 145 items of the TSR (from which the 50 items of the Therapeutic Bond Scales are derived) requires 10–15 minutes to complete
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

Feasibility	
Copyright	1999, APA[1]
Web or scanning options	No details
Training details	No details/NA (patient self-report measure)
Administration/process details	Questionnaire completed after the session (see Acceptability above). In the development study,[1] data were collected following the third or fourth session
Support from measure developers	Authors invite contact for readers interested in converting item scores into corresponding TSR items[1]
FAQ facility	No details
Precision	
Scale type	Three-point Likert-type scale. To make the scale scores directly analogous to the TSR items (which the patient rates on a 0–1–2 scale), each bond scale score was transformed so that it ranged from 0 to 20. Thus, a score of 0 = no experience of that bond dimension, 10 = some experience and 20 = a lot of experience
Normative data	No details
Résumé	
Strengths	On the whole, the reliability and predictive validity of the Therapeutic Bond Scales is supported by the results.[1] The scales are developed from an existing scale (TSR; Orlinsky and Howard, 1966, 1986b)
Weaknesses	Only one primary article was identified specifically on the Therapeutic Bond Scales
Areas for further research	Further validation in other settings and patient groups
Primary reference	
1. Saunders SM, Howard KI, Orlinsky DE. The Therapeutic Bond Scales: psychometric characteristics and relationship to treatment effectiveness. *Psychol Assess* 1989;**1**:323–30.	
Secondary references	
Orlinsky DC, Howard KI. Process and outcome in psychotherapy. In Garfield SL, Bergin AE editors. *Bergin and Garfield's handbook of psychotherapy and behavior change*, 3rd ed. New York: John Wiley; 1986.	
Orlinsky DC, Howard KI. *Therapy session reports. Forms P and I.* Chicago: Institute of Juvenile Research; 1966.	

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

T3
Therapeutic Factors Inventory (TFI)

General details

Author	Lese KP, MacNair-Semands RR
Language	English
Country of publication/development	USA
Publication date	2000
Publisher	NA

Purpose and overview

The TFI was designed to provide a comprehensive, empirically based measure to determine the presence or absence of therapeutic factors in a particular group. The TFI assesses group member perceptions of the degree to which the therapeutic factors described by Yalom (1995) are present in a given group. Scales of the TFI include instillation of hope, universality, imparting information, altruism, corrective re-enactment of the primary family group, development of socialising techniques, imitative behaviour, interpersonal learning, cohesiveness, catharsis and existential factors

Theoretical orientation	Interpersonal
Population details	Participants taken from groups of various formats including open-ended therapy groups, structured groups, support groups, process-orientated experiential and supervision groups[1]
Perspective	Patient
Measure used by	Group therapists
Other versions	None
Notes	

Areas of therapist–patient interaction addressed: Map

Therapy context: type of therapy

Roles: friend/companion; confidant; consumer

Patient engagement: motivation; expectation/preferences; commitment; intentions

Framework: complementary; reciprocal; collaborative/participative/involving; congruent; challenging

Therapeutic techniques: transference; responsiveness/receptiveness/attunement; exploration

Threats to the relationship: defensive; resistance; withdrawal

Outcomes: working alliance; safety/secure base; cohesion; cathartic experience; expression of feelings; corrective emotional experience

Information derived from items

Dimensions

Each scale comprises nine items	Sample items
Altruism	It has impressed me that people in my group can be so kind and giving to one another
Catharsis	I can 'let it all out' in my group
Cohesiveness	We cooperate and work together in group
Corrective re-enactment of primary family group	I have found myself playing the same role in the group that I played in my family at times
Development of socialising techniques	Group helps me learn how to be more clear and direct with other people
Existential factors	In group I have learned that I am responsible for my own improvement
Imitative behavior	I learn how other people act in group and imitate them when it is appropriate

continued

Imparting information	We share ideas about resources in group
Instillation of hope	I don't think the group helps me feel any better about the future
Interpersonal learning	I learn in the group by interacting with the other group members
Universality	We have little in common in my group
Reliability	
The TFI demonstrates adequate internal consistency and adequate test–retest reliability for all but two scales	
Split-half	No details
Internal consistency	Coefficient alphas for the 12 subscales ranged from 0.82 to 0.94
Inter-rater	NA
Test–retest	TFIs were given 1 week apart. Pearson product-moment correlations were used to supply test–retest reliability estimates and were as follows: Altruism: $r = 0.87$ Catharsis: $r = 0.89$ Cohesiveness: $r = 0.93$ Existential factors: $r = 0.64$ Instillation of hope: $r = 0.88$ Imitative behaviour: $r = 0.78$ Imparting information: $r = 0.84$ Interpersonal learning: $r = 0.74$ Corrective re-enactment of primary family group: $r = 0.28$ Development of socialising techniques: $r = 0.72$ Universality: $r = 0.85$ All p were < 0.001 apart from corrective re-enactment of primary family group which was < 0.05
Validity	
Item analysis gave support to the content validity of the TFI The TFI displays partial convergent validity with the IIP; the measure was responsive to differences in interpersonal styles as measured by the IIP Factor analysis demonstrated a lack of independence between the scales	
Face	The therapeutic factors were defined using Yalom's (1995) descriptions and items were generated based on this formulation. The authors (who had graduate-level specialities in group psychotherapy) independently generated the items. They then critiqued and revised items based on lack of clarity, lack of correspondence with the factor definition and redundancy. Following this critique problematic items were revised or eliminated, leaving a total of 174 items
Content	An item analysis was completed. Items with the lowest correlations between that item and the relevant factor score were removed. To ensure an equal number of items per scale, different cut-off levels were used for each scale. Following analysis of the item correlations, 75 items were deleted, leaving a total of 99 items with nine items per scale
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	Construct validity was demonstrated by support for the authors' hypotheses[2] • Therapeutic factors in the group changed over time (average of six group sessions). Analysis of means revealed a significant increase ($p < 0.05$) for universality, instillation of hope, imparting information, recapitulation of the family, cohesiveness and catharsis • Significant correlation patterns with the Inventory of Interpersonal Problems (IIP; Horowitz *et al*., 1988) show that the perceptions of therapeutic factors are related to participants' interpersonal problems

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Convergent	Significant correlations with the IIP revealed: • A relationship between difficulty being submissive and the perception of altruism ($r = -0.42$, $p < 0.005$), socialisation ($r = -0.40$, $p < 0.005$), imitative behaviour ($r = -0.37$, $p < 0.05$) and interpersonal learning ($r = -0.40$, $p < 0.005$). • Positive correlations were found between the perception of altruistic behaviours in the group and having problems related to lack of assertiveness ($r = 0.36$, $p < 0.05$), and being too responsible ($r = 0.30$, $p < 0.05$)
Discriminant	No details
Factor structure	Many of the therapeutic factors scales correlated significantly with one another. The scales that had the weakest correlation with the other scales were imparting information and corrective re-enactment of the primary family group
Responsiveness	
Discriminative (between individuals)	Participants who rated themselves as being overly dominant (according to the IIP) tended to see the group as less altruistic, less apt to promote socialising, having less modelling through imitative behaviour, and giving less interpersonal feedback.[2] See correlation patterns under Construct validity
Evaluative (within individual across time)	Therapeutic factors in the group changed over time (average of six group sessions). Analysis of means revealed a significant increase ($p < 0.05$) for universality, instillation of hope, imparting information, recapitulation of the family, cohesiveness and catharsis[2]
Acceptability	
Number of items	99
Administration method	Self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	2002, Eastern Group Psychotherapy Society
Web or scanning options	No details
Training details	No details
Administration/process details	Forms completed out of session and returned to an anonymous drop box
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Items assessed along a seven-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree). A few items in each scale were reverse-coded to reduce response bias
Normative data	No details

continued

Résumé	
Strengths	Adequate internal consistency. Nine out of 11 scales display adequate test–retest reliability. Promising evidence for Construct validity
Weaknesses	Length of instrument. Lack of interdependence of the factors: the correlated scales could be measuring the same variable due to scale construction
Areas for further research	Further factor analytic work
	Further psychometric research on diverse samples
	The relationship of the TFI to a behavioural measure using external criteria would provide construct validation for the TFI in a multimethod approach

Primary reference

1. Lese KP, MacNair-Semands RR. The Therapeutic Factors Inventory: development of a scale. *Group* 2000;**24**:303–17.

Secondary reference

2. MacNair-Semands RR, Lese KP. Interpersonal problems and the perception of therapeutic factors in group therapy. *Small Group Res* 2000;**31**:158–74.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

T4
Therapist Action Scale (TAS)

General details

Author	Hoyt MF
Language	English
Country of publication/development	USA
Publication date	1981
Publisher	NA

Purpose and overview

To assess the emphasis of specific actions of therapists during dynamic psychotherapy, actions that could be identified and repeated if they were found to be related to successful treatment outcomes

Theoretical orientation	Psychodynamic
Population details	Neurotic-level outpatients with stress response syndrome[1–3]
Perspective	Therapist and rater
Measure used by	Therapists/research therapists
Other versions	Parallel patient action scale
Notes	

Areas of therapist–patient interaction addressed: Map

Therapy context: type of therapy; responsibilities

Roles: expert/authority/leader

Individual differences: expert/authority/leader

Therapist engagement: openness; listening; hope/encouragement

Patient engagement: intentions; expectation/preferences

Framework: collaborative/participative/involving

Non-verbal communication: laughter/humour; silence

Information derived from items

Dimensions

None

Reliability

Adequate inter-rater reliability was demonstrated by the TAS, but the inter-rater reliability between therapist and judge was partial, with agreement on certain items inadequate

Adequate test–retest reliability was demonstrated in one study

Split-half	No details
Internal consistency	No details
Inter-rater	Two raters rated each of 100 sessions, so $N = 200$. Finn's r statistic was used to calculate inter-rater reliability: The median Finn's r for the final list of 25 TAS items is 0.76, range 0.92 to 0.44[1] Between the independent judges the inter-rater reliability was found to be adequate to good, the median Kendall tau coefficient being 0.60 (range 0.82 to 0.19). The median tau coefficient for therapist-rater reliability was 0.33 (range 0.72 to –0.15)[2]

continued

	For those variables that had enough variance to allow meaningful computation, Kendall's tau indicated that nine of the 17 items achieved inter-rater reliability at or beyond a 0.6 coefficient level. Using a less stringent 0.4 cut-off, which Kraemer (1981) suggests for such complex clinical judgements, indicates at least marginally acceptable levels of inter-rater reliability for 14 of the 17 items. The same results are obtained if inter-rater reliability is computed using the ICC[3]
	Therapist-independent judge correlations ranged from 0.77 to –0.15[3]
Test–retest	Five raters repeated six sessions of ratings to $N = 31$. Finn's r statistic was used to calculate test–retest reliability:
	For the TAS median Finn's $r = 0.87$, range 0.97 to 0.68[1]
Validity	
Procedures were carried out to ensure face/content validity of the TAS. The TAS displayed adequate convergent validity with the Patient Action Scale (PAS) and convergent validity with good–poor sessions was demonstrated to a greater degree by judge ratings of the TAS than by therapist ratings. The factor structure provided some support for the authors' hypotheses regarding the clustering of items	
Face	The TAS was derived from previous therapist rating activities plus the authors own theoretical and clinical backgrounds. A number of items were revised and added to earlier forms on the basis of suggestions made by supervisors and advanced psychiatry residents in the authors' psychotherapy study group[1]
Content	Seven parallel TAS items were endorsed less than 20% of the time by at least one judge and were excluded from the final forms of the TAS[1]
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	While each TAS item refers to a discrete, meaningful activity, the evidence regarding the clustering of items (see Factor structure) that theoretically might be expected to go together provides evidence of construct validity to complement the face validity[1]
Convergent	There is a substantial correlation between TAS and the PAS for many of the parallel items, with the median Pearson's $r = 0.76$ (range 0.94 to 0.14)[1]
	Significant correlations were obtained between therapists' judgements of the good–poor quality of sessions and five of 26 TAS variables. The independent judge's good–poor ratings were significantly correlated with his ratings for 11 of 26 TAS items, although two of these significant correlations occurred for items that the judge rarely endorsed as occurring and are thus of dubious reliability[3]
Discriminant	No details
Factor structure	By using a correlational measure of association and an average linkage algorithm (Sokol and Sneath, 1973), three clusters emerged on the TAS: 'reactions to therapist', 'working through the stress event' and 'termination'.[1]
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	25
Administration method	Therapist and rater-completed questionnaire
Time taken to complete	No details
Flesch reading age	No details. Items were written to avoid theoretical jargon and/or the need for complex clinical inferences in favour of specific, operationally defined actions

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1981, *Psychotherapy: Theory, Research and Practice*
Web or scanning options	No details
Training details	Each rater undergoes 12 hours of training before beginning the task
Administration/process details	TAS ratings are made after the session. Provisional ratings done for each third of the session as an aid to recall. After doing the ratings for each segment, the rater then reviews the ratings for the three segments and forms a total rating for the entire session for each TAS item. Since the scales are used for rating audio recordings actions must be audible events to be scored
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	For each item the rater makes a judgement of occurrence or non-occurrence. In addition, a global rating is obtained on a five-point Likert-type scale ranging from 1 (occurred but minor) to 5 (major emphasis)
Normative data	No details
Résumé	
Strengths	Adequate inter-rater reliability between judges. Adequate test–retest reliability
Weaknesses	Poor therapist–judge inter-rater reliability
Areas for further research	Research on improving therapists' self-reports of their actions in psychotherapy

Primary references

1. Hoyt MF, Marmar CR, Horowitz MJ, Alvarez WF. The Therapist Action Scale and the Patient Action Scale: instruments for the assessment of activities during dynamic psychotherapy. *Psychother Theory Res Pract* 1981;**18**:109–16.
2. Hoyt MF, Xenakis SN, Marmar CR, Horowitz MJ. Therapists' actions that influence their perceptions of 'good' psychotherapy sessions. *J Nerv Ment Dis* 1983;**171**:400–4.
3. Xenakis SN, Hoyt MF, Marmar CR, Horowitz MJ. Reliability of self-reports by therapists using the Therapist Action Scale. *Psychother The Res Pract* 1983;**20**:314–20.

Secondary reference

4. Windholz MJ, Weiss DS, Horowitz MJ. An empirical study of the natural history of time-limited psychotherapy for stress response syndromes. *Psychother Theory Res Pract Train* 1985;**22**:547–54.

T5
Therapist Behavior Scale (TBS)

General details	
Authors	Duckro P, George C, Beal DG
Language	English
Country of publication/development	USA
Publication date	1980
Publisher	NA
Purpose and overview	
TBS is a specialised research instrument particularly useful for scientists studying the effects of clients' expectations and/or preferences on psychotherapeutic process or outcome. It is designed to assess clients' attitudes regarding highly directive vs not very directive therapists. The initial development of the TBS is reported in a dissertation by Reiter (1967)[2]	
Theoretical orientation	Not specified
Population details	Undergraduate students (172 psychology students;[1] 86 psychology students[2])
Perspective	Patient
Measure used by	Psychotherapist (secondary level/clinic)
Other versions	Ref. 1 details a modified version from the original dissertation[2]
Notes	TBS records preferences for therapist behaviour. In the primary articles identified, university students place themselves in the client's position and rate what type of psychotherapist behaviour they would prefer[1]
Areas of therapist–patient interaction addressed: Map	
Therapy context	
Roles	
Therapist engagement	
Patient engagement	
Framework	
Information derived from items	
Dimensions	
High directiveness	In the original version[2] high directiveness comprised 18 items and low directiveness 14 items. Eight items failed to correlate with the total score and were eliminated as scored items
Low directiveness	In the modified version[1] 14 items scored as high directive, 13 as low directive and 13 were buffer items
Reliability	
Reliability information is relatively sparse. Test–retest reliability was adequate	
Split-half	No details
Internal consistency	"Results from factor analysis show the scale as an internally consistent measure"[1] (see Factor structure below)
Inter-rater	NA
Test–retest	Modified version: 0.72 in pilot study over a 3-week period[3] Original version: 0.79 over a 3-week period[2]
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Validity	
Validity information is relatively sparse. Factor analyses supported the existence of two dimensions (high directive and low directive)	
Face	No details
Content	Five clinical psychologists independently judged whether items from a pool adequately reflected high and low directiveness in therapists' behaviour. At least four of the five judges agreed in their evaluation of 40 items from the pool; these items comprised the original scale[2]
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	An R-type factor analysis was conducted using the principal components method for factor extraction. The two main factors (high directiveness and low directiveness) accounted for 20.6% of the variance. A Q-type factor analysis examined the differential patterns of response on the high and low directive items. 48 subjects were taken from the original sample (the 24 highest scorers and the 24 lowest). Results showed it may be more useful to score the high and low directive factors separately before obtaining the overall score[1]
Responsiveness	
Discriminative (between individuals)	See Factor structure
Evaluative (within individual across time)	No details
Acceptability	
Number of ttems	40
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1980, *Catalog of Selected Documents in Psychology*
Web or scanning options	No details
Training details	No details
Administration/process details	No details
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Five-point ordinal, Likert-type scale,[1] originally a dichotomous (agree–disagree) response option[2]
Normative data	No details

continued

Notes

The TBS has been used in a study examining the hypothesis that failure to meet client preferences for high or low directive counsellor style would adversely affect interpersonal processes. 48 university students participated in a counselling intervention analogue orientated around their actual problems. There was no evidence that failure to meet client preference adversely affected interview process[4]

Résumé

Strengths	Seems useful for measuring preferences for or expectation of therapist directiveness in psychotherapy. Test–retest reliability is adequate and factor analyses support the existence of two factors (high and low directiveness)
Weaknesses	Validated on students imagining themselves in the client's position rather than using real clients. On the whole, reliability and validity information was sparse
Areas for further research	Further reliability and validity testing, and validation of the measure with clients who have actually received or are receiving therapy

Primary references

1. Duckro P, George C, Beal DG. Internal structure of the therapist behavior scale. *Catalog of Selected Documents in Psychology* 1980; 10 MS. 2139, p. 91.
2. Reiter M. Variables associated with the degree of preferred directiveness in therapy. *Dissert Abst* 1967;**27**(3679B).

Secondary references

3. Duckro P, George C, Beal DG. Malleability of preference for therapists' response style. *Psychol Rep* 1978;**43**:299–304.
4. Duckro PN, George CE. Effects of failure to meet client preference in a counseling interview analogue. *J Counsel Psychol* 1979;**26**:9–14.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

T6
Therapist Representation Inventory (TRI) – Fourth Section: Record of Dreams

General details	
Authors	Geller JD, Cooley RS, Hartley D
Language	English
Country of publication/development	USA
Publication date	1982
Publisher	NA
Purpose and overview	
This is the final of four measures that comprise the Therapist Representation Inventory (TRI), which is a means of examining patients' representations of their therapists. This section is concerned with the content, vividness and frequency of patients' dreams that feature their therapist, the purpose being to examine the consistency of representations between waking and sleeping states	
Theoretical orientation	Psychodynamic
Population details	Professional psychotherapists as current or past psychoanalysis or psychotherapy patients
Perspective	Self-report
Measure used by	Researchers
Other versions	No details
Notes	As part of the TRI, the initial standardisation of the Free Response Task was based on data from 206 professional psychotherapists with experience as psychoanalysis or psychotherapy patients. The participants had 1–46 years of clinical experience, were aged from 25–75 years, 120 were male and 66 were currently in therapy[1,2] Although professional therapists do not constitute a typical sample of patients, they were enlisted because of their presumed ability to access and describe representations of their therapists[2]
Areas of therapist–patient interaction addressed: Map	
The measure (a record of dreams) could address any area of therapist–patient interaction depending upon the content of each individual client's dream and method of analysis	
Dimensions	
Dimensions are specific to the analytic method. Readers are referred to ref. 3 for a detailed report of a thematic analysis with a sample drawn from the 206 psychotherapists employed in the standardisation of the TRI	
Reliability	
Reliability is an issue for the method of analysis rather than for the measure itself. Readers are referred to ref. 3 for a detailed report of a thematic analysis. The sample was drawn from the 206 psychotherapists employed in the standardisation of the TRI and the report includes results of reliability assessments of the themes identified	
Split half	NA
Internal consistency	NA
Inter-rater	NA
Test–retest	NA

continued

Validity	
The measure has face validity in that patients report their own dreams directly. As with reliability, validity is more applicable to the method of analysis than to the measure itself. Readers are referred to ref. 3 for a report of a thematic analysis of dreams. The participants were drawn from the 206 psychotherapists employed in the standardisation of the TRI	
Face	Patients freely record their own dreams
Content	NA
Criterion (a) concurrent	NA
Criterion (b) predictive	NA
Construct	NA
Convergent	NA
Discriminant	NA
Factor structure	NA
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	NA
Administration method	No details
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1982, Baywood Publishing Co
Web or scanning options	No details
Training details	No details
Administration/process details	Patients are asked to rate the vividness and frequency of dreams in which their therapist appears, and to report such a dream
Support from measure developers	Copies of TRI may be obtained from Jesse D. Geller, Department of Psychology, Yale University, Box 11A Yale Station, New Haven, CT 06520, USA[2]
FAQ facility	No details
Precision	
Scale type	Qualitative data
Normative data	No details
Résumé	
Strengths	The measure allows clients to describe their experiences (dreams) in their own words
Weaknesses	The measure may only be useful to psychodynamically orientated services
Areas for further research	Further assessment of psychometric properties, including independent work with more diverse client groups
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Primary references

1. Geller JD, Cooley RS, Hartley D. Images of the psychotherapist: a theoretical and methodological perspective. *Imagin Cognit Personal* 1982;**1**:123–46.
2. Orlinsky DE, Geller JD, Tarragona M, Farber B. Patients' representations of psychotherapy: a new focus for psychodynamic research. *J Consult Clin Psychol* 1993;**61**:596–610.

Secondary references

3. Rhode AB, Geller JD, Farber BA. Dreams about the therapist: mood, interactions, and themes. *Psychotherapy* 1992;**29**:536–44.

T7
Therapist Representation Inventory (TRI) – Free Response Task

General details	
Authors	Geller JD, Cooley RS, Hartley D
Language	English
Country of publication/development	USA
Publication date	1982
Publisher	NA
Purpose and overview	
The Free Response Task is one of four measures that comprise the Therapist Representation Inventory (TRI), which is a means of examining patients' representations of their therapists. The Free Response Task is designed to evaluate the thematic content and conceptual level of patients' representations of their therapist. The task for patients is to write an open-ended description of their therapists	
Theoretical orientation	Psychodynamic
Population details	Professional psychotherapists as current or past psychoanalysis or psychotherapy patients
Perspective	Self-report
Measure used by	Researchers
Other versions	No details
Notes	As part of the TPI, the initial standardisation of the Free Response Task was based on data from 206 professional psychotherapists with experience as psychoanalysis or psychotherapy patients. The participants had 1–46 years of clinical experience, were aged from 25 to 75 years, 120 were male and 66 were currently in therapy[2] Although professional therapists do not constitute a typical sample of patients, they were enlisted because of their presumed ability to access and describe representations of their therapists[2]
Areas of therapist–patient interaction addressed: Map	
The Free Response Task could address any area of therapist–patient interaction that is relevant to the respondent	
Dimensions	
The dimensions that are tapped depend upon the content of each individual representation and the method of analysis. The researchers in this study used the system devised by Blatt *et al*. to score the 'conceptual level' of object representations[1,2]	
Reliability	
The reliability of the Free Response Task is a matter for the scoring or analytic method, rather than for the task itself	
Split-half	No details
Internal consistency	No details
Inter-rater	No details
Test–retest	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Validity	
The Free Response Task has content validity as it enables clients to describe their own representations of their therapist in their own words. As reliability, validity is matter of the scoring or analytic method	
Face	No details
Content	Clients describe their therapist in their own words
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	NA
Evaluative (within individual across time)	No details
Acceptability	
Number of items	NA
Administration method	Open-ended questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1982, Baywood Publishing Co
Web or scanning options	No details
Training details	No details
Administration/process details	At the top of an otherwise blank piece of paper, participants are instructed as follows: 'Please describe your (current/previous) therapist. Take no longer than five minutes to complete this task'[1,2]
Support from measure developers	Copies of TRI may be obtained from Jesse D. Geller, Department of Psychology, Yale University, Box 11A Yale Station, New Haven, CT 06520, USA[2]
FAQ facility	No details
Precision	
Scale type	Qualitative data
Normative data	No details
Résumé	
Strengths	The measure allows clients to describe their therapist in their own words
Weaknesses	The measure may only be useful to psychodynamically orientated services
Areas for further research	Further assessment of psychometric properties, including independent work with more diverse client groups

continued

Primary references
1. Geller JD, Cooley RS, Hartley D. Images of the psychotherapist: a theoretical and methodological perspective. *Imagin Cognit Personal* 1982;**1**:123–46. 2. Orlinsky DE, Geller JD, Tarragona M, Farber B. Patients' representations of psychotherapy: a new focus for psychodynamic research. *J Consult Clin Psychol* 1993;**61**:596–610.
Secondary references
None

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

T8
Therapist Representation Inventory (TRI) – Therapist Embodiment Scale (TES)

General details	
Authors	Geller JD, Cooley RS, Hartley D
Language	English
Country of publication/development	USA
Publication date	1982
Publisher	NA

Purpose and overview	
The Therapist Embodiment Scale (TES) is one of four measures that comprise the Therapist Representation Inventory (TRI). The TES is designed to provide information regarding the formal properties (as distinct from thematic content) of patients' representations of therapy and their therapists	
Theoretical orientation	Psychodynamic
Population details	Professional psychotherapists with experience as current or past patients of psychoanalysis or psychotherapy
Perspective	Self-report
Measure used by	Researchers
Other versions	No details
Notes	As part of the TPI, the initial standardisation of the TES was based on data from 206 professional psychotherapists with experience as psychoanalysis or psychotherapy patients. The participants had 1–46 years of clinical experience, were aged from 25 to 75 years, 120 were male and 66 were currently in therapy[2] Although professional therapists do not constitute a typical sample of patients, they were enlisted because of their presumed ability to access and describe representations of their therapists[2]

Areas of therapist–patient interaction addressed: Map

Therapy context: type of therapy

Roles: attachment figure

Individual differences: attachment styles

Therapist engagement: empathy/sensitivity

Therapeutic techniques: responsiveness/receptiveness/attunement

Non-verbal communication: touch

The therapist-client interaction information is derived from the scale items in refs 1 and 2

Dimensions	
Visualisation	Visualisation is defined by four items (1, 2, 5 and 6), which are mainly visual images of the therapist in their office
Proximal embodiment	Proximal embodiment is defined by three items (primarily item 8, 'I experience in myself certain characteristic bodily sensations', as well as items 11 and 12). Proximal embodiment involves a blend of imagery derived from more immediate kinaesthetic, proprioceptive and tactile senses
Conversational–conceptual	The conversational–conceptual dimension was tentatively identified. It has four items (4, 6, 7 and 10) and includes mainly auditory and lexical representations of real and imagined conversations with the therapist (e.g. item 7, 'I think of my therapist as making specific statements to me'). There are no details of the total number of items that make up this dimension

continued

Reliability	
Three scales based on the dimensions of the TES (see Dimensions and Factor structure) were assessed for internal consistency with the participants described above. The reliability coefficients for the scales based on the visualisation, proximal embodiment and conversational–conceptual dimensions were adequate, partial and inadequate, respectively[2]	
Significant part–whole correlations for all 12 TES items indicate that they can also be scored as an overall index of representational vividness[1,2]	
Split-half	No details
Internal consistency	Coefficients for the TES dimensions were 0.72 (visualisation), 0.69 (proximal embodiment) and 0.49 (conversational–conceptual)[2]
	Item-total correlations were significant ($p < 0.001$) for all 12 TES items[1,2]
Inter-rater	No details
Test–retest	No details
Validity	
The correlation between TES scores and client-rated outcome was significant but too low to establish predictive validity. The visualisation (factor 1) also failed to demonstrate adequate predictive validity when correlated with client-rated improvement	
With data from the participants described above, principal components factor analysis with varimax rotation suggested three dimensions: visualisation (four items), proximal embodiment (three items) and conversational–conceptual (four items)[1,2]	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	The correlation between client-rated improvement and (a) TES total scores was 0.22 ($p < 0.005$), and (b) visualisation (factor 1) was 0.15, ($p < 0.04$)[1]
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	Three factors or dimensions emerged from a factor analysis of the data:
	Visualisation: four items with factor loadings of 0.45 to 0.75[1,2]
	Proximal embodiment: three factors with factor loadings from 0.48 to 0.93[1,2]
	Conversational–conceptual: four factors with factor loadings from 0.32 to 0.51[1,2]
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	12
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Feasibility	
Copyright	1982, Baywood Publishing Co
Web or scanning options	No details
Training details	No details
Administration/process details	Each of the 12 TES items is a statement. Patients respond to each item on a nine-point Likert-type scale anchored at 1 (not al all characteristic) and 9 (highly characteristic)
Support from measure developers	Copies of TRI may be obtained from Jesse D. Geller, Department of Psychology, Yale University, Box 11A Yale Station, New Haven, CT 06520, USA[2]
FAQ facility	No details
Precision	
Scale type	Likert-type with nine points anchored at 1 (not at all characteristic) and 9 (highly characteristic)
Normative data	No details
Résumé	
Strengths	The TES is short with only 12 items. The visualisation dimension has adequate internal consistency with a coefficient of 0.72 and proximal embodiment's internal consistency is just short of adequate with a coefficient of 0.69[2]
Weaknesses	The measure may only be useful to psychodynamically orientated services
Areas for further research	The conversational–conceptual dimension does not have adequate internal consistency (coefficient = 0.49)[2]
	The TES did not establish adequate predictive validity when correlated with client-rated improvement. Coefficients were 0.22 ($p < 0.005$) (TES total scores) and 0.15 ($p < 0.04$) (visualisation)[1]
	Further assessment of psychometric properties, including independent work with more diverse client groups
Primary references	
1. Geller JD, Cooley RS, Hartley D. Images of the psychotherapist: a theoretical and methodological perspective. *Imagin Cognit Personal* 1982;**1**:123–46. 2. Orlinsky DE, Geller JD, Tarragona M, Farber B. Patients' representations of psychotherapy: a new focus for psychodynamic research. *J Consult Clin Psychol* 1993;**61**:596–610.	
Secondary references	
None	

T9
Therapist Representation Inventory (TRI) – Therapist Involvement Scale (TIS)

General details	
Authors	Geller JD, Cooley RS, Hartley D
Language	English
Country of publication/development	USA
Publication date	1982
Publisher	NA
Purpose and overview	
The Therapist Involvement Scale (TIS) is one of four measures that comprise the Therapist Representation Inventory (TRI), which is a means of examining patients' representations of their therapists. The TIS is designed to examine the functional themes that characterise patients' thoughts, wishes and fantasies about their therapist	
Theoretical orientation	Psychodynamic
Population details	Professional psychotherapists as current or past psychoanalysis or psychotherapy patients
Perspective	Self-report
Measure used by	Not specified
Other versions	No details
Notes	As part of the TPI, the initial standardisation of the TIS was based on data from 206 professional psychotherapists with experience as psychoanalysis or psychotherapy patients. The participants had 1–46 years of clinical experience, were aged from 25 to 75 years, 120 were male and 66 were currently in therapy[2] Although professional therapists do not constitute a typical sample of patients, they were enlisted because of their presumed ability to access and describe representations of their therapists[2]
Areas of therapist–patient interaction addressed: Map	
Therapy context: type of therapy; boundaries; responsibilities Roles: attachment figure; expert/authority/leader Individual differences: social support; attachment styles; defensive style/repression Therapist engagement: genuineness; praise/affirmation Patient engagement: motivation; expectation/preferences; attraction; commitment; intentions Framework: collaborative/participative/involving Non-verbal communication: touch Threats to the relationship: sexual involvement; resistance; withdrawal Outcomes: compliance; satisfaction; working alliance: affective bond; goals; expression of feelings The therapist–client interaction information is derived from the scale's items	

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Dimensions	
	Each of the six dimensions represents aspects of interaction between the patient and therapist, which the patient continues with mental representations[2]
Continuing the therapeutic dialogue	Consists of four items (11, 16, 26 and 28). Endorsing the items indicates that the patient uses the representation as a means of sustaining the work of therapy in the physical absence of the therapist
Sexual and aggressive involvement	Consists of five items (3, 20, 27, 29 and 35). The content of the items converges on fantasies of physical contact with the therapist, particularly of a sexual and physical nature
The wish for reciprocity	Consists of four items (12, 31, 32 and 34). The items indicate the wish-fulfilling fantasy of an extratherapeutic relationship with the therapist
Failure of benign internalisation	Consists of five items (2, 4, 6, 22 and 36). The items pertain to various issues which may prevent the internalisation of the therapist as a benignly influential other
The effort to create a stable representation of the therapist	Consists of three items (9, 10 and 23). The items indicate a preoccupation with the therapist in an apparent effort to hold on to the therapist in his or her absence
Desiring contact (mourning)	Consists of three items (1, 5 and 38). The items signify mourning the loss of the therapist either between sessions or after therapy has ended
Reliability	
Internal consistency has been found to be adequate for five of the six subscales, and partial for the other[1,2]	
Split-half	No details
Internal consistency	The coefficients (p values not reported) for the six subscales were: Continuing the therapeutic dialogue: 0.86[2] Sexual and aggressive involvement: 0.84[2] The wish for reciprocity: 0.81[2] Failure of benign internalisation: 0.70[2] The effort to create a stable representation of the therapist: 0.67[2] Desiring contact (mourning): 0.76[2]
Inter-rater	No details
Test–retest	No details
Validity	
With data from the sample described above, factor analysis identified six distinguishable dimensions. These dimensions reflect patient–therapist interactions, which patients continue with mental representations of their therapists (see Dimensions and Internal consistency for more details)[2]	
Continuing the therapeutic dialogue demonstrated partial predictive validity when correlated with client-perceived beneficial outcome of therapy. The other five factors did not significantly correlate with this outcome measure[1]	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	Continuing the therapeutic dialogue significantly correlated with perceived beneficial outcome ($r = 0.33$, $p < 0.001$)[2]
Construct	No details
Convergent	No details
Discriminant	No details

continued

Factor structure	Factor analysis identified six dimensions: Sexual and aggressive involvement has six items whose factor loadings range from 0.34 to 0.83[1,2] The wish for reciprocity has nine items whose factor loadings range from 0.33 to 0.74[1,2] Continuing the therapeutic dialogue has four items whose factor loadings range from 0.62 to 0.67[1,2] Failures of benign internalisation has eight items whose factor loadings range from 0.36 to 0.61[1,2] The effort to create a stable representation of the therapist (initially labelled the effort to create a therapist introject) has five items whose factor loadings range from 0.36 to 0.59[1,2] Desiring contact (mourning) has five items whose factor loadings range from 0.36 to 0.66[1,2]
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	38
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1982, Baywood Publishing Co
Web or scanning options	No details
Training details	No details
Administration/process details	Each of the 38 TES items is a statement, e.g. 'I miss my therapist'. Patients respond to each item on a nine-point Likert-type scale anchored at 1 (not at all characteristic) and 9 (highly characteristic)
Support from measure developers	Copies of TRI may be obtained from Jesse D. Geller, Department of Psychology, Yale University, Box 11A Yale Station, New Haven, CT 06520, USA[2]
FAQ facility	No details
Precision	
Scale type	Likert-type scale anchored at 1 (not at all characteristic) and 9 (highly characteristic)
Normative data	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Résumé

Strengths	Five of the six dimensions have adequate internal consistency (coefficients range from 0.76 to 0.86), and the effort to create a stable representation of the therapist is not far short with a coefficient of 0.67[2]
Weaknesses	The scale is fairly lengthy with 38 items. It is also limited because it is for use in psychodynamic therapy only. The scale failed to predict client-perceived beneficial outcome of therapy. The dimension continuing the therapeutic dialogue has a highly significant, but only partially adequate relationship, while the other five dimensions showed no significant relationship to the outcome measure[1]
Areas for further research	Further assessment of psychometric properties, including independent work with more diverse client groups

Primary references

1. Geller JD, Cooley RS, Hartley D. Images of the psychotherapist: a theoretical and methodological perspective. *Imagin Cognit Personal* 1982;**1**:123–46.
2. Orlinsky DE, Geller JD, Tarragona M, Farber B. Patients' representations of psychotherapy: a new focus for psychodynamic research. *J Consult and Clin Psychol* 1993;**61**:596–610.

Secondary references

None

T10
Truax and Carkhuff (1967) Scales

General details	
Authors	Truax CB, Carkhuff RR
Language	English
Country of publication/development	USA
Publication date	1967
Publisher	NA
Purpose and overview	
The purpose of the scales is to measure the facilitative conditions of therapy. The scales are: therapist's accurate empathy (AE), non-possessive warmth and genuineness. There is one scale for each condition	
Theoretical orientation	Counselling psychology
Population details	Clinical adults, clinical adolescents. Late adolescents, students and adults with interpersonal and/or vocational difficulties
Perspective	Independent rater
Measure used by	Researchers, and therapists for training and supervision
Other versions	Five-point rating scale (see ref. 3)
Notes	
Areas of therapist–patient interaction addressed: Map	
Roles: advocate; expert/authority/leader; good object	
Therapist engagement: empathy/sensitivity; warmth; genuineness; respect; praise/affirmation	
Framework: convergent; congruent; controlling; directive; exploration	
Therapeutic techniques: responsiveness/receptiveness/attunement; feedback	
Threats to the relationship: critical; defensive	
Therapist-patient interaction information derived from descriptions of the scales	
Truax and Carkhuff (1967) Scales	
Accurate empathy	Defined by Truax as "the therapist's sensitivity to current feelings and verbal facility to communicate this understanding in a language attuned to the client's current feelings" (see ref. 8) The scale is designed to define the stages of accurate empathy. At a high level of accurate empathy, the message 'I am with you' is clear. At a low level, the therapist may go off on a tangent of their own or misinterpret what the client is saying
Non-possessive warmth	The scale is designed to define the stages of non-possessive warmth. At a low level the therapist is actively offering advice or giving clear negative regard. At a high level the therapist communicates warmth without restriction. There is a deep respect for the client's worth as a person and their rights as a free individual
Genuineness	The scale is designed to define the stages of genuineness or self-congruence. The scale begins from a low level where the therapist presents a façade to a high level where the therapist is freely and deeply himself/herself

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability

Non-possessive warmth and genuineness have demonstrated adequate inter-rater reliability.[4] Inter-rater reliability of AE has been shown to be partial in one study[8] and adequate in two[3,4]

Split-half	No details
Internal consistency	No details
Inter-rater	Raters are trained to meet a criterion of an inter-rater reliability of 0.50[6]
	With two raters over nine ratings, the AE scale achieved inter-rater reliabilities ranging from 0.80 to 0.98 (*p* not reported)[3]
	With five clients, inter-rater reliabilities (Pearson's *r*) for AE, non-possessive warmth and genuineness were 0.85, 0.88 and 0.86, respectively[4]
	Inter-rater reliability of the AE, using the Spearman–Brown formula, was 0.64[8]
	Using a procedure suggested by Ebel (1951; see ref. 2), intraclass reliability coefficients were 0.92 and 0.95 for two rating groups[2]
Test–retest	Raters are trained to meet a criterion of a test–retest reliability of 0.50[6]

Validity

AE demonstrated adequate convergent validity when correlated, using Pearson product-moment, with the Carkhuff (1969, see ref. 3) five-point version of the empathy scale. Applying a z-score transformation also revealed no significant differences between the two scales[3]

AE, non-possessive warmth and genuineness have demonstrated adequate convergent validity when correlated with each other, and with Shapiros' corresponding scales (understanding–not understanding, accepting–rejecting and genuine–false).[6] It has been argued that these interscale correlations are evidence of poor validity of the scales, as they were designed to measure distinct concepts and therefore should not converge too highly[5]

AE has shown partial to adequate convergent validity with therapist statements regarding client emotions[8]

Convergent validity was not established between the Truax and Carkhuff (1967) scales and the Barrett-Lennard Relationship Inventory scales (no correlation was significant at 0.05)[4]

Non-possessive warmth and genuineness were correlated (at therapist and session levels) with each other and individually with another measure of empathy and measures of immediacy and self-disclosure. The resulting coefficients varied to include from not significant, to partial and adequate[1]

AE correlations with therapist statements about facts and therapy procedures were not adequate to establish discriminant[3]

Face	The scales are derived from Truax's (1961) Group Process Scales, which are derived from Rogers' (1957) necessary and sufficient conditions
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	The 1967 nine-point AE scale was correlated with the Carkhuff (1969; see ref. 7) five-point version. The correlation between the two was 0.89 ($n = 42$, $p < 0.001$)[3]
	Using a *z*-score transformation, no significant differences were found between the 1967 nine-point and the Carkhuff (1969; see ref. 3) five-point accurate empathy scales[3]
	The scales were individually correlated (Pearson product-moment) with corresponding scales, resulting in the following coefficients:
	AE with understanding–not understanding: 0.67 ($p < 0.01$)[6]
	Non-possessive warmth with accepting–rejecting: 0.89 ($p < 0.01$)[6]
	Genuineness with genuine–false: 0.78 ($p < 0.01$)[6]

continued

	Interscale correlations resulted in the following coefficients: AE – non-possessive warmth: 0.58 ($p < 0.01$)[6] AE – genuineness: 0.53 ($p < 0.01$)[6] Non-possessive warmth–genuineness: 0.73 ($p < 0.01$)[6] The AE scale was correlated, using Pearson product-moment, with 22 other measures. The most significant correlations were with: The therapist making statements in a specific manner about the client's emotion: 0.52 ($p = 0.011$)[8] The proportion of total therapist responses consisting of statements about emotion: 0.38, 0.52 ($p = 0.01$)[8] The proportion of total responses in which emotion is mentioned: 0.36 ($p = 0.01$)[8] Pearson product-moment correlations were calculated between the Truax and Carkhuff Scales and the four Barrett-Lennard Relationship Inventory scales (empathic understanding, level of regard, unconditionality of regard and congruence). No significant relationships were found ($p > 0.05$)[4] Using therapist means, the non-possessive warmth and genuineness scales were correlated with each other and individually correlated with three other measures of facilitative conditions [empathy (Bergin and Solomon; see ref. 1), immediacy of relationship and facilitative self-disclosure (Carkhuff, 1969; see ref. 1)]. Pearson product-moment coefficients ($n = 15$, all significant at 0.05 or 0.01) were as follows: Non-possessive warmth: 0.85 (genuineness); 0.93 (empathy); 0.85 (immediacy); and 0.70 (self-disclosure)[1] Genuineness: –0.85 (empathy); 0.87 (immediacy); and 0.87 (self-disclosure)[1] The same correlations were calculated for session means. Correlations were calculated for high ($n = 25$) and low ($n = 26$) empathic therapist groups, Coefficients for high and low groups, respectively, were as follows: Non-possessive warmth: 0.52 and 0.78, $p < 0.01$ (genuineness); 0.49, $p < 0.05$ and 0.53, $p < 0.01$ (empathy); ns and 0.41, $p < 0.05$ (immediacy); ns and 0.44, $p < 0.05$ (self-disclosure)[1] Genuineness: 0.43, $p < 0.05$ and 0.51, $p < 0.01$ (empathy); 0.43 and 0.45, $p < 0.05$ (immediacy); 0.60, $p < 0.01$ and ns (self-disclosure)[1]
Discriminant	There is controversy regarding the validity of the scales with regard to their convergence with each other and with other scales measuring different constructs (e.g. AE–genuineness, AE–self-disclosure). The debate centres around whether therefore some of the correlation coefficients are too high given that the scales were designed as distinct constructs[1,5,7] AE correlated negatively with therapist questions about facts, (–0.27, $p = 0.05$) and with therapist statements about therapy procedures (–0.29, $p = 0.05$)[8]
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	There was a trend for non-possessive warmth and genuineness (facilitative conditions) to be more closely related to each other than to action-orientated measures (e.g. self-disclosure). This trend was stronger with a stronger trend in high facilitative vs low facilitative therapists, although there are no details regarding the statistical significance of the difference between the two groups[6]
Evaluative (within individual across time)	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Acceptability	
Number of items	NA
Administration method	Rating scale
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1967, Aldine, Chicago
Web or scanning options	No details
Training details	Raters work with training manuals and are trained to meet inter-rater and test–retest reliability criteria of 0.50[6]
Administration/process details	Therapists' responses from segments of recorded therapy sessions are rated
Support from measure developers	The rating training manuals include an introduction to the concepts being rated, specific descriptions of the scale points and examples of therapist behaviour for each scale point[6]
FAQ facility	No details
Precision	
Scale type	Likert-type. Nine-point scales with lower scores indicating lesser degrees of the therapist quality
Normative data	No details

Notes

Other research uses of the Truax and Carkhuff (1967) scales include:

A comparison of three methods of assessing the psychotherapist's empathy: (a) the Accurate Empathy Scale, (b) the Conjunctive Empathy Scale, and (c) the Raskin Empathy Scale[9]

A study of sources of variance in 'accurate empathy' ratings[10]

A study of the influence of counsellor empathy, student sex and grade level on perceived counsellor role[11]

An investigation of relationship between theoretical orientation and therapists' empathy, warmth and genuineness[12]

A study of the level of empathy displayed by both members of an interactive dyad as influenced by the A–B types of the individuals of the dyad[13]

An investigation into some relationships between the verbal behavior of 20 experienced therapists and their therapist effectiveness[14]

An investigation into the relationship between counselling systems stages and counsellor effectiveness as measured by the Carkhuff Empathy Scale[15]

An examination of the counsellor's skills when counselling clients with sexual problems[16]

A study of the relationship between a counsellor's own experience with a problem and his or her empathy in responding to a client with the same problem[17]

An examination of the importance of therapist–client agreement for therapy outcome in group therapy[18]

A comparison of complete novices and experienced professionals on interviewer behaviour and efficacy[19]

An examination of personality similarities in the patient–therapist dyad to determine whether these similarities were conducive to the formation of a positive therapeutic relationship[20]

A preliminary study of the Accurate Empathy scale[21]

A study of the effectiveness of in Parent Effectiveness Training (PET)[23]

continued

An investigation of whether empathy increases with clinical experience[24]

A study of treatment outcome as a function of the therapist's focus on the patient's source of anxiety and accurate empathy[25]

A study designed to clarify the relationship between certain characteristics of therapist's responses and patient intrapersonal exploration, and certain characteristics of group atmosphere and patient self-exploration[26]

A study evaluating therapeutic conditions through group psychotherapy[27]

A study of the relationship between ratings on the Truax Accurate Empathy Scale and linguistic variables of therapist speech[28]

Résumé

Strengths	The scales are derived from Rogers' necessary and sufficient conditions
	Training is required for raters, which, while being a limitation because it is time-consuming, is thorough and improves reliability. Raters are trained to meet inter-rater and test–retest reliabilities of 0.50, and the rating training manuals include an introduction to the concepts being rated, specific descriptions of the scale points and examples of therapist behaviour for each scale point[6]
	Inter-rater reliabilities for non-possessive warmth and genuineness were 0.88 and 0.86. respectively.[4] Accurate empathy was also shown to have adequate inter-rater reliability in two studies (Pearson's *r* ranged from 0.80 to 0.98[3] and 0.85[4]) and partial inter-rater reliability in another (Spearman–Brown formula was 0.64[8])
	Intraclass reliability coefficients were 0.92 and 0.95 for two rating groups[2]
	The scales have also demonstrated convergent validity. When AE was correlated with the Carkhuff empathy scale, the coefficient was 0.89 ($p < 0.001$).[3] AE also converged with Shapiro's understanding–not understanding ($r = 0.67$, $p < 0.01$). Non-possessive warmth and genuineness adequately converged with Shapiro's equivalent accepting–rejecting and genuine–false scales ($r = 0.89$ and 0.78 respectively, $p < 0.01$)[6]
	Non-possessive warmth and genuineness demonstrated convergent validity when correlated with three other measures of facilitative conditions [empathy (Bergin and Solomon; see ref. 1), immediacy of relationship and facilitative self-disclosure (Carkhuff, 1969, see ref. 1)]
	Correlation coefficients ranged from 0.70 to 0.93 (non-possessive warmth) and from 0.85 to 0.87 (genuineness)[1]
Weaknesses	Training is required for raters, which is time-consuming (although it also has advantages, which are discussed above)
Areas for further research	Further assessment of psychometric properties, e.g. criterion and construct validity

Primary references

1. Barrow JC. Interdependence of scales for the facilitative conditions: three types of correlational data. *J Consult Clin Psychol* 1977;**45**:654–59.
2. Blaas CD, Heck EJ. Accuracy of accurate empathy ratings. *J Counsel Psychol* 1975;**22**:243–6.
3. Engram BE, Vandergoot D. Correlation between the Truax and Carkhuff scales for measurement of empathy. *J Counsel Psychol* 1978;**25**:349–51.
4. McWhirter JJ. Two measures of the facilitative conditions. *J Counsel Psychol* 1973;**20**:317–20.
5. Rappaport J, Chinsky JM. Accurate empathy: confusion of a construct. *Psychol Bull* 1972;**77**:400–4.
6. Shapiro DA. Relationships between expert and neophyte ratings of therapeutic conditions. *J Consult Clin Psychol* 1968;**32**:87–9.
7. Truax CB. The meaning and reliability of accurate empathy ratings: a rejoinder. *Psychol Bull* 1972;**77**:397–9.
8. Wenegrat A. A factor analytic study of the Truax Accurate Empathy Scale. *Psychother Theory Res Pract* 1974;**11**:48–51.

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Secondary references

9. Bachrach H, Mintz J, Luborsky L. On rating empathy and other psychotherapy variables: an experience with the effects of training. *J Consul Clin Psychol* 1971;**36**:445.
10. Beutler LE, Johnson DT, Neville CW, Workman SN. Some sources of variance in 'accurate empathy' ratings. *J Consult Clin Psychol* 1973;**40**:167–9.
11. Conklin RC, Nakoneshny M. The influence of counselor empathy, student sex, and grade level on perceived counselor role. *Can Counsel* 1973;**7**:206–12.
12. Fischer J, Paveza GJ, Kickertz NS, Hubbard LJ, Grayston SB. The relationship between theoretical orientation and therapists' empathy, warmth, and genuineness. *J Counsel Psychol* 1975;**22**:399–403.
13. Gillam S, McGinley H. AB similarity-complementarity and accurate empathy. *J Clin Psychol* 1983;**39**:512–19.
14. Hayden B. Verbal and therapeutic styles of experienced therapists who differ in peer-rated therapist effectiveness. *J Counsel Psychol* 1975;**22**:384–89.
15. Lutwak N, Hennessy JJ. Conceptual systems functioning as a mediating factor in the development of counseling skills. *J Counsel Psychol* 1982;**29**:256–60.
16. McConnell LG. An examination of the counselor's skills when counselling clients with sexual problems. *Fam Coord* 1976;**25**:183–8
17. Neher LA, Dicken C. Empathy and the counsellor's experience of the client's problem. *Psychother Theory Res Pract* 1975;**12**:360–4.
18. Peake TH. Therapist–patient agreement and outcome in group therapy. *J Clin Psychol* 1979;**35**:637–46.
19. Pope B, Nudler S, Vonkorff MR, McGhee JP. The experienced professional interviewer versus the complete novice. *J Consult Clin Psychol* 1974;**42**:680–90.
20. Rinaldi RC. Patient–therapist personality similarity and the therapeutic relationship. *Psychother Private Pract* 1987;**5**:11–29.
21. Shapiro DA. The rating of psychotherapeutic empathy: a preliminary study. *Br J Soc Clin Psychol* 1970;**9**:148–51.
22. Shapiro DA. Naive British judgments of therapeutic conditions. *Br J Soc Clin Psychol* 1973;**12**:289–94.
23. Therrien ME. Evaluating empathy skill training for parents. *Soc Work* 1979;**24**:417–19.
24. Thomson DH, Hassenkamp A-M, Mansbridge C. The measurement of empathy in a clinical and a non-clinical setting. Does empathy increase with clinical experience? *Physiotherapy* 1997;**83**:173–80.
25. Traux CB, Wittmer J. The effects of therapist focus on patient anxiety source and the interaction with therapist level of accurate empathy. *J Clin Psychol* 1971;**27**:297–9.
26. Truax CB, Carkhuff RR, Kodman F. Relationships between therapist-offered conditions and patient change in group psychotherapy. *J Clin Psychol* 1965;**21**:327–9.
27. Truax CB. The process of group psychotherapy: relationship between hypothesized therapeutic conditions and intrapersonal exploration. *Psychol Monogr* 1961;**75**(7):35.
28. Wenegrat A. Linguistic variables of therapist speech and accurate empathy ratings. *Psychother Theory Res Pract* 1976;**13**:30–3.

V1
Vanderbilt Negative Indicators Scale (VNIS)

General details	
Author	Strupp H
Language	English
Country of publication/development	USA
Publication date	1980
Publisher	NA
Purpose and overview	
The VNIS measures characteristics of the patient, therapist and their interaction, which may lead to negative change	
Theoretical orientation	Various, although the scale is anchored in psychodynamic conceptions
Population details	Patients from the Vanderbilt Project were used
Perspective	Independent rater. Rated by clinicians
Measure used by	Practitioners
Other versions	Vanderbilt Negative Indicators Scale – Short
Notes	No details
Areas of therapist–patient interaction addressed: Map	
Individual differences: level of functioning	
Patient engagement: motivation, expectations/preferences	
Framework: controlling; collaborative/participative/involving; flexible/rigid	
Threats to the relationship: critical; resistance	
Outcomes: achieving a working relationship	
Inferred from representative items listed in ref. 1[1]	
Dimensions	
Subscales:	
Patient personal qualities	17 items, e.g. problems with verbal self-expression
Therapist personal qualities	Nine items, e.g. lack of respect for the patient
Errors in technique	Ten items, e.g. destructive interventions
Patient–therapist interaction	Two items, e.g. problems in the therapeutic relationship
Global session ratings	Four items, e.g. dull interaction
Reliability	
VNIS is deliberately sensitive to deficiencies and errors in a therapist's performance, yet there is considerable room for disagreement on what constitutes good as well as poor practice. VNIS calls for value judgements. Reliabilities vary between good and low for the different subscales	
Split-half	No details
Internal consistency	Alpha coefficients. Interpreter reliabilities for subscales range from 0.26 (patient–therapist interaction) to 0.81 (therapist personal qualities), suggesting that some subscales do not tap unified dimensions (Sandell, 1981)[1]
Inter-rater	Pearson's *R*. Generally good for subscales, although coefficients low for errors in technique ($r = 0.58$) and patient–therapist interaction ($r = 0.63$) (Sandell, 1981)[1]
Test–retest	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Validity	
Present and past studies have demonstrated some predictive validity, but this is the only area of validity, to have been addressed	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	The comparison of VNIS total score between high and low outcome cases was found to be statistically significant, suggesting predictive validity (Strupp *et al*. 1980, presentation)
	Correlation coefficients between the VNIS subscale scores and a composite measure (outcome measure – overall improvement) for each of the first three sessions was calculated. All subclasses except therapist personal qualities demonstrated at least one significant correlation with outcome. Errors in technique showed the strongest and most consistent relationship to outcome[1]
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individuals)	No details
Acceptability	
Number of items	42
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	No details
Web or scanning options	No details
Training details	Raters should ideally be well-trained therapists who are familiar with the range of therapeutic practices and the quality of performance[1]
Administration/process details	No details
Support from measure developers	Manual available
FAQ facility	No details
Precision	
Scale type	The first judgement involves a simple dichotomous decision as to whether a given characteristic is present or absent. Items judged ‘present’ are rated on an ordinal Likert scale from 1 to 5, reflecting the frequency or intensity of the negative indicator
Normative data	No details

continued

Résumé

Strengths	Some predictive validity has been demonstrated
Weaknesses	The reliability of the VNIS is subject to raters' value judgements. Reliable use requires the raters to be therapeutically experienced and well trained
	The different subscales vary in their reliability from low to good
Areas for further research	Only predictive validity has been addressed; future work should aim to cover other areas of validity

Primary reference

1. Suh CS, Strupp HH, O'Malley SS. The Vanderbilt process measures: the Psychotherapy Process Scale (VPPS) and the Negative Indicators Scale (VNIS). In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford Press; 1986; pp. 285–323.

Secondary references

2. Sachs JS. Negative factors in brief psychotherapy: an empirical assessment. *J Consult Clin Psychol* 1983;**51**:557–64.
3. Eaton TT, Abeles N, Gutfreund MJ. Negative indicators, therapeutic alliance, and therapy outcome. *Psychother Res* 1993;**3**:115–23.
4. Raytek HS, McCrady BS, Epstein EE, Hirshch LS. Therapeutic alliance and the retention of couples in conjoint alcoholism treatment. *Addict Behav* 1999;**24**:317–30.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

V2
Vanderbilt Negative Indicators Scale – Short (VNIS-S)

General details	
Author	Nergaard MO
Language	English
Country of publication/development	USA/Norway
Publication date	1989
Publisher	No details
Purpose and overview	
Measures characteristics of the patient, therapist and their interaction, which may lead to negative change (same as for the VNIS, but Nergaard shortened the scale for this particular study)	
Theoretical orientation	Psychodynamically orientated therapy
Population details	See below
Perspective	Independently rated by psychologists
Measure used by	Therapists
Other versions	Unshortened Vanderbilt Negative Indicators Scale
Notes	*Clients:* Age range 20–80 (mean age 55.5). All patients met the criteria: (1) history of positive interpersonal relationships; (2) no evidence of organic brain syndrome or mental deficiency; (3) no evidence of serious substance abuse; (4) no evidence of suicidal or homicidal potential. All patients suffered from neurotic or character disorders or both *Practitioners:* 11 male, six female. Ages ranging from 31 to 75. Each had at least five years' postdoctoral experience, including some training in brief dynamic psychotherapy *Raters:* One male, one female. Both held PhDs in clinical psychology. Both were trained in a psychodynamic framework
Areas of therapist–patient interaction addressed: Map	
Therapy context: responsibilities	
Patient engagement: motivation	
Therapist engagement	
Threats to the relationship	
Outcomes	
Inferred from brief description of the scale and a couple of example items	
Dimensions	
Subscales	No dimension information but subscales
Patient	17 items
Global session	Number of items not specified
Therapist–patient interaction	Number of items not specified

continued

Reliability	
The internal consistency of the VNIS-S was partial for the patient and global sessions subscales at session 1, and by session 8 their respective reliabilities were adequate. The internal consistency of the interaction subscale was inadequate at both sessions 1 and 8	
The subscales had partial to adequate correlation with one another	
Inter-rater reliability ranged from being inadequate to adequate, but was generally high	
Split-half	No details
Internal consistency	Alpha coefficient of 0.84. Overall internal reliabilities for the subscales were higher for session 8 (S8) than session 1 (S1). Reliabilities for the subscales were 0.64, 0.61 and 0.43 for the patient, global session and interaction subscales, respectively, at S1. At S8, their respective reliabilities were 0.78, 0.73 and 0.54
	Correlations between patient and interaction were 0.64 (S1) and 0.72 (S8), between patient and global session were 0.67 (S1) and 0.62 (S8), and between interaction and global session were 0.72 (S1) and 0.62 (S8)
Inter-rater	Generally high, with the exception of item 21 (dull interaction). Range from 0.45 to 0.98, with an average coefficient alpha of 0.84
Test–retest	No details
Validity	
Nergaard's Guilt scale was a consistently better predictor of outcome than was the VNIS-S and had more correlations with outcome measures. Other areas of validity need to be addressed	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	VNIS-S scores were predictive only in the therapist ratings of outcome
Construct	No details
Convergent	Correlations between the Vanderbilt subscales and the outcome measures (Symptom Checklist 90-R, Global Assessment Scale, Overall Change Rating) are mostly negative. Seven of these 24 correlations are significant at the $p < 0.05$ level
Discriminant	No details
Factor structure	No details
Responsiveness	
Discriminative (between individuals)	NA
Evaluative (within individual across time)	NA
Acceptability	
Number of items	No details
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Feasibility	
Copyright	Nergaard (1989)
Web or scanning options	No details
Training details	No details
Administration/process details	No details
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert
Normative data	No details
Résumé	
Strengths	Adequate internal consistency for the patient and global subscales, adequate correlation between subscales and generally high inter-rater reliability
Weaknesses	Poor internal consistency of the Interaction subscale, relatively poor predictive validity
Areas for further research	More areas of validity need to be addressed
Primary reference	
1. Nergaard MO, Silberschatz G. The effects of shame, guilt, and the negative reaction in brief dynamic psychotherapy. *Psychotherapy* 1989;**26**:330–7.	
Secondary references	
None	

V3
Vanderbilt Psychotherapy Process Scale – 80 item (VPPS-80)

General details	
Author	Strupp H
Language	English
Country of publication/development	USA
Publication date	1983
Publisher	NA
Purpose and overview	
A general purpose instrument designed to assess the positive and negative aspects of the patient's and the therapist's behaviour and attitude that are expected to facilitate or impede therapy, and their interaction which may be related to outcome. The overall purpose is to provide meaningful indices of the therapeutic process, which may be used in comparative analyses or studied in relation to pre- or post-therapy assessments made by patients, therapists or independent clinicians	
Theoretical orientation	Various/range. Intended to be largely neutral with respect to any particular theory of psychotherapy, and to be applicable to a wide range of therapeutic interventions. Rational emotive therapy,[7] analytical psychotherapy,[8] time-limited dynamic psychotherapy,[9,15] time-limited, interpretive individual psychotherapy,[12] manual-guided psychotherapy, Short-Term Interpersonal Psychotherapy of Depression (IPT),[13] brief (16-session) psychodynamic therapies,[17] personal construct psychotherapy and rationalist cognitive therapy[18]
Population details	Clinical adults/school-aged children
Perspective	Independent rater. Raters should be uninvolved, external observers either from the actual therapy sessions or from video- or audiotapes of therapy. Raters should have some knowledge of the therapy process, yet graduate students with minimal clinical experience can also use the instrument reliably[3]
Measure used by	Research, clinical practice and training[3,6]
Other versions	Vanderbilt Psychotherapy Process Scale – 44 item Intake version (35 items)
Notes	*Clients:* Caucasian therapist–client dyads (mean age 30)[4] School age[7] 25 unmarried male college students with elevated scores on MMPI Scales 2, 7 and 0 who were participating in a psychotherapy outcome study[8] 21- and 30-year-old women suffering from anxiety and relationship difficulties[11] 41 adult daughters or daughters-in-law caring for a frail person living in the community[14] Adult outpatients[17] *Practitioners:* Therapist[2] Non-professional ('inherently helpful' college professors) therapists[8] Four peer counsellors and four professional counselors[14] *Raters:* Recent PhD clinical psychologists[3] Two advanced clinical psychology graduate students[8]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Areas of therapist–patient interaction addressed: Map	
Therapy context: influence; power/coercion; values	
Roles: confidant; expert/authority/leader; advocate	
Individual differences: level of functioning; attachment styles; defensive style/repression	
Therapist engagement: hope/encouragement; praise/affirmation; empathy/sensitivity; warmth; listening; openness; respect	
Patient engagement: motivation; commitment; intentions; expectations/preferences	
Framework: collaborative/participative/involving; focused; flexible/rigid; controlling	
Therapeutic techniques: exploration; responsiveness/receptiveness/attunement	
Nonverbal communication: paralinguistics	
Threats to the relationship: defensive; hostility/anger; withdrawal; fear; resistance; critical	
Emotional expression: cathartic experience; expression of feelings	
Changing view of self with others	
Inferred from items as fully listed in manual	
Dimensions	
Subscales:	
Patient participation	Eight items: Patient's active involvement in the therapy interaction
Therapist hostility	Six items. Level of negativism, hostility or distrust displayed by the patient
Patient psychic distress	Nine items: Level of emotional distress and feelings of discouragement expressed by the patient
Patient exploration	Seven items: patient's level of self-examination and exploration of feelings and experiences
Patient dependency	Six items: Patient's reliance and dependency on the therapist
Therapist exploration	13 items: Therapist's attempts to examine the psychodynamics underlying the patient's problems
Therapist warmth and friendliness	Nine items: therapist's display of warmth and emotional involvement with the patient
Negative therapist attitude	Six items. Therapist's attitudes that might intimidate or threaten the patient
	Items: the first three items are used to obtain global impressions regarding the quality of the relationship, the overall productivity of the session and the patient's level of functioning. Remaining items are divided into two sections, patient and therapist items (40 and 37 items, respectively). Each section comprises two parts, one dealing with characteristics of patient's 'behaviour' during the session, and the second consisting of adjectives which describe each patient's 'demeanour'.[3] There are three overall ratings
Reliability	
Adequate inter-rater reliability and internal consistency for the VPPS-80 have been demonstrated	
Split-half	No details
Internal consistency	Ranges from 0.81 to 0.96 (median 0.92)[1]
Inter-rater	Ranges from 0.79 to 0.94 (median 0.92)[1]
	The level of inference required for ratings was minimised for the purpose of enhancing inter-rater reliability (e.g. Kiesler, 1973; Strupp, 1960)[3]
Test–retest	No details

continued

Validity	
Predictive validity for the VPPS-80 ranges from partial to adequate. Factor analysis supports the eight subscales of the VPPS-80	
Face	No details
Content	No details
Criterion (a) concurrent	No details
Criterion (b) predictive	Suh and O'Malley (1982) examined the relationship between change scores on patient participation and outcome. The correlation coefficients are high (ranging from 0.43 to 0.72, $p < 0.05$ to $p < 0.001$), suggesting that the amount of change on patient qualities tapped by this scale can be seen as a meaningful index of outcome[2]
	Multiple regression analyses: The dimension 'patient involvement' showed the most consistent relationship with outcome, predicting overall improvement and improvement in target complaints made from the three perspectives (patient, therapist, clinician, F ranging from 3.34 to 6.35, all $p < 0.05$ or < 0.01). The dimension 'exploratory processes' predicted the therapists' ratings of both overall improvement ($F = 5.35$, $p < 0.01$) and in-target complaints ($F = 3.07$, $p < 0.05$). The dimension of 'therapist offered relationship' predicted only the therapist's overall rating of improvement ($F = 4.36$, $p < 0.05$)[3]
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	Principal components factor analysis with varimax rotation led to the following seven factors: patient participation, patient hostility, therapist warmth and friendliness, negative therapist attitude, patient exploration, therapist exploration and patient psychic distress[1]
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	80
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1983, American Psychological Association
Web or scanning options	No details
Training details	Judges undergo training to a criterion level. The judges continued to rate additional tapes until they reached the criterion level of inter-rater reliability ($r = 0.90$)[1]
Administration/process details	A systematic sampling method was used (5 minutes from the beginning, middle and end of hour) from the first three sessions[1]
Support from measure developers	Manual available which defines each item
FAQ facility	No details
	continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Precision	
Scale type	Ordinal, Likert. Ratings are made on a standard scale ranging from 1 (not at all) to 5 (great deal)
Normative data	No details

Notes

Five media forms were evaluated: (1) transcript; (2) audiotape; (3) videotape; (4) audiotape supplemented with a transcript; (5) videotape plus transcript. It was found that transcripts were generally inadequate for making VPPS ratings, particularly if minimal training is provided or raters with low levels of clinical experience are used. Audio or videotapes are preferable[1]

Résumé	
Strengths	Adequate inter-rater reliability and internal consistency for the VPPS-80 have been demonstrated
	Predictive validity for the VPPS-80 ranges from partial to adequate. Factor analysis supports the eight subscales of the VPPS-80
	VPPS is simple, robust and meaningful, despite lacking ability to assess precisely each patient and therapist interaction[2]
Weaknesses	Primary articles only address a limited number of psychometric properties
	Transcripts are generally inadequate for making VPPS ratings, particularly if minimal training is provided or raters with low levels of clinical experience are used[1]
Areas for further research	Further investigation of psychometric properties

Primary references

1. O'Malley SS, Suh CS, Strupp HH. The Vanderbilt Psychotherapy Process Scale: a report on the scale development and a process-outcome study. *J Consult Clin Psychol* 1983;**51**:581–86.
2. Suh CS, O'Malley SS, Strupp HH, Johnson ME. The Vanderbilt Psychotherapy Process Scale (VPPS). *J Cogn Psychother* 1989;**3**:123–54.
3. Suh CS, Strupp HH, O'Malley SS. The Vanderbilt process measures: the Psychother Proc Scale (VPPS) and the Negative Indicators Scale (VNIS). In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford Press; 1986: pp. 285–323.

Secondary references

4. Bachelor A, Salame R. Participants' perceptions of dimensions of the therapeutic alliance over the course of therapy. *J Psychother Pract Res* 2000;**9**:39–53.
5. Baer J. Evaluating practice: assessment of the therapeutic process. *J Soc Work Educ* 2001;**37**:127–36.
6. Borders LD, Fong ML. Evaluations of supervisees: brief commentary and research report. *Clin Supervis* 1991;**9**:43–51.
7. Flanagan R, Povall L, Dellino M, Byrne L. A comparison of problem solving with and without rational emotive behavior therapy to improve children's social skills. *J Ration Emot Cogn Behav Ther* 1998;**16**:125–34.
8. Gomes-Schwartz B, Schwartz JM. Psychotherapy process variables distinguishing the inherently helpful person from the professional psychotherapist. *J Consult Clin Psychol* 1978;**46**:196–7.
9. Henry WP, Butler SF, Strupp HH, Schacht TE, Binder JL. Effects of training in time-limited dynamic psychotherapy: changes in therapist behavior. *J Consult Clin Psychol* 1993;**61**:434–40.
10. Henry WP, Strupp HH. The therapeutic alliance as interpersonal process. In Horvath AO, Greenberg LS, editors. *The working alliance: Theory, research, and practice. Wiley series on personality processes*. New York: Wiley; 1994. pp. 51–84.
11. Holland SJ, Roberts NE, Messer SB. Reliability and validity of the Rutgers Psychotherapy Progress Scale. *Psychother Res* 1998;**8**:104–10.
12. Piper WE, Ogrodniczuk JS, Joyce AS, McCallum M, Rosie JS, O'Kelly JG, *et al*. Prediction of dropping out in time-limited, interpretive individual psychotherapy. *Psychotherapy* 1999;**36**:114–22.
13. Rounsaville BJ, Chevron ES, Prusoff BA, Elkin I, Imber S, Sotsky S, Watkins J. The relation between specific and general dimensions of the psychotherapy process in interpersonal psychotherapy of depression. *J Consult Clin Psychol* 1987;**55**:379–84.
14. Smith MF, Tobin SS, Toseland RW. Therapeutic processes in professional and peer counseling of family caregivers of frail elderly people. *Social Work* 1992;**37**:345–51.

continued

15. Soldz S, Budman S, Demby A. The relationship between main actor behaviours and treatment outcome in group psychotherapy. *Psychother Res* 1992;**2**:52–62.
16. Strupp HH. *Vanderbilt Psychotherapy Process Scales (VPPS) rater manual*. Nashville, TN: Vanderbilt University; 1983.
17. Windholz MJ, Silberschatz G. Vanderbilt Psychotherapy Process Scale: a replication with adult outpatients. *J Consult Clin Psychol* 1988;**56**:56–60.
18. Winter DA, Watson S. Personal construct psychotherapy and the cognitive therapies: different in theory but can they be differentiated in practice? *J Construct Psychol* 1999;**12**:1–22.
19. Wiseman H, Shefler G, Caneti L, Ronen Y. A systematic comparison of two cases in Mann's time-limited psychotherapy: An events approach. *Psychother Res* 1993;**3**:227–44.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

V4
Vanderbilt Therapeutic Alliance Scale (VTAS)

General details	
Author	Hartley DE
Language	English
Country of publication/development	No details
Publication date	1983
Publisher	Erlbaum Press
Purpose and overview	
The instrument attributes a successful therapeutic alliance to the presence or absence of six factors: positive climate, therapist intrusiveness, client resistance or anxiety (Langs, 1976), client motivation (Greenson, 1967) and client responsibility (Bordin, 1979)	
Theoretical orientation	VTAS represents a theoretical blend of dynamic and eclectic frameworks (ref. 2, p. 264) e.g. person-centred,[4] behavioural,[4] psychoanalytic,[4] time-limited therapy condition or a time-unlimited therapy condition,[7] and couple therapy for alcoholism[8]
Population details	Clinical adults. See below for details
Perspective	Trained judges rate each item (ideally two working alongside each other, to ensure reliability). Clinician observers[3]
Measure used by	Practitioners and researchers
Other versions	No details
Notes	*Clients:* Participants' reported mean days of cocaine use in the 30 days before entering treatment was 12.98 (SD 8.83)[1] Data drawn from psychotherapy sessions that were part of a randomised clinical trial evaluating treatment for co-morbid cocaine and alcohol dependence. All male (28)[2] Age 18–25 years, college students, single, suffered from anxiety, depression and discomfort relating to peers (females)[3] Data for this study were taken from eight cases of brief (12–20 sessions) psychotherapy[4] Inpatients (seven men and seven women) from a psychiatric unit[6] Students at a university-based counselling centre[7] *Practitioners:* Psychologists[1–3] Clinicians[1,2] Psychiatrists[1] Psychotherapists[3] Four male and four female therapists, ranging in age from 34–78 years, with 5–42 years of postdoctoral experience[4] Therapists, whose experience ranged from 15 years of postdoctoral practice to a 1-year predoctoral practicum[8] *Raters:* Six doctoral students in counselling or clinical psychology served as raters for the observer-rated working alliance measures[4]

continued

Areas of therapist–patient interaction addressed: Map	
Therapy context: power/coercion	
Roles: expert/authority/leader	
Therapist engagement: hope/encouragement; empathy/sensitivity; genuineness; respect; support/tolerance; openness	
Patient engagement: motivation; commitment	
Framework: focused; challenging; reciprocal; convergent; structuring	
Therapeutic techniques: responsiveness/receptiveness/attunement; exploration	
Non-verbal communication: silence	
Threats to the relationship: hostility/anger; defensive; critical; intrusive	
Outcomes: changing view of self with others; general satisfaction; achieving a working relationship	
Inferred from the items of the scale	
Dimensions	
	There are 44 items in the questionnaire (14 relating to the patient, 18 relating to the therapist and 12 to their interaction)
Positive climate	20 items (mainly therapist items, a few patient and interaction), e.g. hopeful (therapist), feels supported (patient)
Patient resistance	Three patient and two interaction items, e.g. hostile (patient), power struggle (interaction)
Therapist intrusiveness	Four therapist items: e.g. therapist imposes own values
Patient motivation	Two patient and four interaction items, e.g. desire to overcome (patient), focus on task (interaction)
Patient responsibility	Six patient items, e.g. carries out tasks
Patient anxiety	One patient item: anxious; one interaction item: awkward pauses
Reliability	
Adequate internal consistency has been demonstrated for the VTAS. Partial to adequate inter-rater reliability has been demonstrated. No other reliability areas have been addressed in the primary articles	
Split-half	No details
Internal consistency	0.96.[1] Patient scale (0.93), therapist scale (0.84), together (0.93), 0.95 full scale alpha (Hartley and Strupp, 1983), 0.93 (Tichenor *et al*., 1989), coefficient alpha for the full scale of 0.87 (Carroll *et al*., 1997)[1]
	0.95 alpha values. Scores from both judges were combined into composite measures. 0.95 represents the total scale score internal reliability, while for the subscales, alpha values were 0.92 (for therapist subscale) 0.89 (for patient subscale), and 0.87 (for patient–therapist subscale)[3]
	For the three scales, coefficient alpha ranged from 0.87 to 0.92. Overall internal consistency was 0.93[4]
Inter-rater	0.68 fixed ICC, 0.70 random ICC, 0.69 mean ICC (Hartley and Strupp 1983), 0.74 (Tichenor *et al*., 1989), ICC for full scale was 0.59 (Carroll *et al*., 1997)[1]
	For the eight sessions that were rated by all raters, 0.6 was the random-effect ICC estimate[2]
	0.97 product-moment correlation. This value applied for all ratings across all sessions and all items. (The raters agreed exactly on 49% of the ratings, and were discrepant by only one 'step' on another 43%)[3]
	For the three scales, inter-rated reliability ranged from 0.79 to 0.90, using the Ebel's R[4]
Test–retest	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Validity	
VTAS has at least partial concurrent validity with the Working Alliance Inventory – Observer (WAI-O), Working Alliance Inventory – Therapist (WAI-T), California Psychotherapy Alliance Scales (CALPAS) and Penn Helping Alliance Rating Scale (Penn)	
The predictive validity of the VTAS is unclear owing to different implications resulting from different studies and methods of calculation	
Convergent validity is demonstrated with adequate correlations between VTAS, CALPAS, Penn and WAI-O	
Face	No details
Content	No details
Criterion (a) concurrent	VTAS total correlated 0.75 ($p < 0.001$) with WAI-O, 0.24 with WAI-C, 0.30 with WAI-T, 0.52 ($p < 0.001$) with CALPAS and 0.47 ($p < 0.001$) with Penn[1]
Criterion (b) predictive	Correlations between alliance and outcome were 0.49 ($p < 0.001$) for all treatments, 0.46 ($p < 0.001$) for cognitive-behavioural therapy (CBT) and 0.55 ($p < 0.001$) for Twelve-Step Facilitation (TSF)[2]
	Association was calculated (using ANOVAs) between measures of therapeutic alliance for each case (VTAS scores) and three outcome groups: 'dropout' attended five or fewer of their 25 sessions, 'low outcome' and 'high outcome' (those who completed their sessions). Therapeutic alliance scores did not vary significantly between these three outcome groups ($p > 0.05$)[3]
Construct	No details
Convergent	Pearson correlation of VTAS with: WAI-O 0.87 ($p < 0.001$), WAI-C 0.02 (ns) and WAI-T 0.36 (ns)[2]
	The VTAS correlated with CALPAS 0.80 ($p < 0.01$), Penn 0.51 (ns), WAI-O 0.84 ($p < 0.01$), WAI-C 0.13 (ns) and WAI-T 0.09 (ns)[4]
Discriminant	No details
Factor structure	Principal components analysis was conducted (type of rotation not stated). Bartlett's test was used to determine the number of significant factors and six factors were extracted, 'borrowing' items across all three subscales of the questionnaire (patient, therapist and interaction)[3]
Responsiveness	
Discriminative (between individuals)	Conducted a one-way, three-level MANOVA and found that on the VTAS, mean alliance ratings were significantly lower ($p < 0.05$) in the clinical management (in comparison to CBT, TSF and each of aforementioned treatments plus disulfiram) group[1]
Evaluative (within individual across time)	No details
Acceptability	
Number of items	44
Administration method	Questionnaire and interview
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1983, Erlbaum Press
Web or scanning options	No details

continued

Training details	The raters completed a training course using tapes published by the American Academy of Psychotherapy. The training procedure called for raters to listen to 15 minutes of an interview and make independent ratings, followed by comparison and discussion on items on which there was disagreement
Administration/process details	Listen to taped interviews. The first, middle and last 5-minute segments from each session were chosen to provide the raters with the best overview of the entire session. In order to sample across time in these cases, the sessions at the quartile points were examined. For dropout cases, all sessions were rated. At the end of each session (15 minutes), the raters were instructed to complete a rating form
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert. The items are scored on a six-point Likert scale ranging from 0 (none at all) to 5 (a great deal) to reflect the extent to which the rater observed the behaviours
Normative data	No details
Résumé	
Strengths	Adequate internal consistency and convergent validity have been demonstrated for the VTAS. Partial to adequate inter-rater reliability has been demonstrated and the VTAS has at least partial concurrent validity
	Results suggest some discriminative validity of the VTAS
Weaknesses	No other reliability areas have been addressed in the primary articles. The predictive validity of the VTAS is unclear
Areas for further research	Future studies should address more areas of reliability and attempt to clarify predictive validity findings

Primary references

1. Cecero JJ, Fenton LR, Frankforter TL, Nich C, Carroll KM. Focus on therapeutic alliance: the psychometric properties of six measures across three treatments. *Psychotherapy* 2001;**38**:1–11.
2. Fenton LR, Cecero JJ, Nich C, Frankforter TL, Carroll KM. Perspective is everything: the predictive validity working alliance instruments. *J Psychother Pract Res* 2001;**10**:262–8.
3. Hartley DE, Strupp HH. The therapeutic alliance: its relationship to outcome in brief psychotherapy. In Masling J, editor. *Empirical Studies of Psychoanalytic Theories*. Hillsdale, NJ: Analytic Press; 1983. pp. 1–27.
4. Tichenor V, Hill CE. A comparison of six measures of working alliance. *Psychotherapy* 1989;**26**:195–9.

Secondary references

5. Henry WP, Strupp HH. The therapeutic alliance as interpersonal process. In Horvath AO, Greenberg LS, editors. *The working alliance: theory, research, and practice. Wiley series on personality processes*. New York: Wiley; 1994. pp. 51–84.
6. Jurich J, Richardson L. The patient's experience of the relationship scale as a means of studying the inpatient psychotherapy relationship. *Psychother Res* 2001;**11**:473–82.
7. Kamin DJ, Garske JP, Sawyer PK, Rawson JC. Effects of explicit time-limits on the initial therapeutic alliance. *Psychol Rep* 1993;**72**:443–48.
8. Raytek HS, McCrady BS, Epstein EE, Hirshch LS. Therapeutic alliance and the retention of couples in conjoint alcoholism treatment. *Addict Behav* 1999;**24**:317–30.
9. Tang TZ, DeRubeis RJ. Sudden gains and critical sessions in CBT for depression. *J Consult Clin Psychol* 1999;**67**:894–904.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

W1
Working Alliance Inventory – Client (WAI-C)

General details	
Authors	Horvath AO, Greenberg L
Language	English
Country of publication/development	USA
Publication date	1986
Publisher	NA
Purpose and overview	
To assess the strengths and dimensions of the alliance as conceptualised by Bordin[12]	
Theoretical orientation	Pan-theoretical
Population details	Adults
Perspective	Client
Measure used by	Therapists/counsellors/research clinicians
Other versions	Short Version (12 items) (available in client, observer and therapist versions) Working Alliance Inventory – Therapist Working Alliance Inventory – Observer
Notes	
Areas of therapist–patient interaction addressed: Map	
Therapy context: influence; power/coercion; responsibilities	
Roles: friend/companion; attachment figure; confidant; expert/authority/leader; protector	
Therapist engagement: all components	
Patient engagement: all components	
Framework: all components	
Therapeutic techniques: responsiveness/receptiveness/attunement; exploration; ruptures/repair	
Threats to the relationship: defensive; critical; fear; resistance; confrontations; withdrawal	
Outcomes: compliance; satisfaction; working alliance; cohesion	
Information derived from items	
Dimensions	
Goal agreement	12 items. The extent to which a client and therapist agree on the goals that are the target of the intervention
Task agreement	12 items. The extent to which a client and therapist agree on the in-counselling behaviours and cognitions that form the substance of the counselling relationship
Bond development	12 items. The extent to which a client and therapist possess mutual trust, acceptance and confidence

continued

Reliability	
The WAI-C total scale and dimensions have adequate internal consistency as measured by Cronbach's alpha. The WAI-C total scale and dimensions have adequate test–retest reliability at 3 weeks.	
Split-half	No details
Internal consistency	Client WAI total scale: 0.95,[1] 0.93,[10] 0.94,[2] 0.93,[6,7] 0.94,[2] 0.93[4]
	Client WAI dimensions: range 0.83 to 0.91,[1] 0.85 to 0.92,[6,7] 0.90 to 0.92,[9] range 0.77 to 0.89[2]
Inter-rater	NA
Test–retest	Client WAI total scale at 3 weeks: 0.80[2]
	Client WAI dimensions: range 0.66 to 0.74[2]
Validity	
Extensive validity research has been carried out on the WAI-C showing the measure to have adequate concurrent and convergent validity and partial predictive validity	
Face	No details
Content	The instrument was generated through a series of sequential ratings and evaluations of prospective items. The initial pool of items was generated on the basis of a content analysis of Bordin's descriptions of the working alliance. The items were rated by experts and professionals. See refs 5 and 6 for details of the item rating summary
Criterion (a) concurrent	For concurrent validity of WAI-C and the WAI-C-S, standardised regression coefficients ranged from 0.80 to 0.97 ($p < 0.01$)[1]
	After excluding data from clients who unilaterally left therapy, the standardised regression coefficient for the final client (WAI-C) and therapist (WAI-T) ratings was 0.33 ($p < 0.01$)[1]
	In a multitrait–multimethod matrix the WAI-C scales correlated with the empathy subscale of the Barrett-Lennard Relationship Inventory (range 0.62 to 0.83)[5,6]
Criterion (b) predictive	The WAI-C fourth-session scores had standardised slope coefficients of 0.36 ($p < 0.01$) for the standardised and averaged 'improvement' score (client and therapist target complaints – improvements and post therapy questionnaire indices) and 0.14 for the standardised and averaged 'symptom' score (client and therapist residual target complaints – severity and client residual global severity Index indices)[1]
	Correlations between alliance and outcome for the WAI-C were not significant[2]
	The total WAI score was correlated significantly ($r = 0.42$, $p < 0.05$) with the total score of the client post therapy questionnaire[5,6]
	The WAI administered in the early stages of therapy was predictive of outcome as measured by therapist ratings of outcome ($r = 0.50$, $p < 0.05$) and patient ratings of outcome ($r = 0.64$, $p < 0.001$)[7]
Construct	No details
Convergent	WAI-C was correlated 0.31 (ns) with California Psychotherapy Alliance Scales–Rater (CALPAS-R), 0.36 (ns) with Penn Helping Alliance Scales (Penn) and 0.02 (ns) with Vanderbilt Therapeutic Alliance Scale (VTAS)[3]
	WAI-C was correlated 0.09 (ns) with WAI–Observer (WAI-O) and 0.43 ($p < 0.01$) with WAI–Therapist (WAI-T)[3]
	WAI-C was correlated 0.33 (ns) with CALPAS-R, 0.02 (ns) with Penn and 0.13 (ns) with VTAS[10]
	WAI-C was correlated –0.18 (ns) with WAI-O and 0.09 (ns) with WAI-T[10]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

	WAI-C was correlated 0.34 (ns) with CALPAS-R, 0.32 (ns) with Penn and 0.24 (ns) with VTAS[2]
	WAI-C was correlated 0.21 (ns) with WAI-O and 0.37 (ns) with WAI-T[2]
	WAI-C was correlated 0.85 ($p < 0.0001$) with CALPAS–Patient (CALPAS-P) and 0.74 ($p < 0.0001$) with Penn Helping Alliance Questionnaire (HAQ)[5]
	WAI-C scales correlated strongly with the bond partnership and confidence scales of the client version of the ARM (ARM–C) at the dyad level (all correlations in 0.80s and 0.90s)[9]
	WAI-C scales correlated strongly with the bond, partnership and confidence scales of the ARM-C, *r* range 0.54 to 0.70[9]
	WAI-C scales correlated less strongly with the ARM scales of the therapist version of the instrument. See ref. 9 for full breakdowns of the results
Discriminant	There is some support for the discriminant validity of the goal scale of the WAI-C as demonstrated by the multitrait–multimethod matrix[6,7]
Factor structure	Random regression coefficients for associations among the subscales ranged from 0.71 to 0.92 ($p < 0.01$)[1]
	The WAI-C comprises three scales of task, bond and goal. There is substantial overlap among the scales; although the authors have managed to separate the goal and bond scales, the task and goal dimensions share a high degree of covariance. However, the authors found that significant portions of outcome variance that were significantly correlated with the goal or bond scales after the outcome variance explained by the dominant task scale had been removed[6,7]
	A confirmatory factor analysis[11] found the WAI (Therapist and Client versions) to measure one general alliance factor as well as the three specific alliance dimensions of task, bond and goal. The validity of the three specific aspects of the alliance was more limited than that of the general alliance factor
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	The WAI was applied across multiple sessions and session-level deviation scores were calculated to provide a measure of session-to-session variation[9]
Acceptability	
Number of items	36
Administration method	Self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1986, Guilford, New York
Web or scanning options	No details
Training details	No details
Administration/process details	Assessed after third counselling session in pilot study,[6,7] but may be completed after any counselling/therapy session
Support from measure developers	No details
FAQ facility	No details

continued

Precision	
Scale type	Each of the three subscales has six positive and six negative items rated on a seven-point Likert scale ranging from 1 (never) to 7 (always). Subscale scores range from 12 to 84 and the total score ranges from 36 to 252. Higher scores reflect more positive ratings of the working alliance
Normative data	Available in ref. 2
Résumé	
Strengths	Adequate convergent validity. High internal consistency and inter-rater reliability. Requires minimal training
Weaknesses	Length of instrument
Areas for further research	Use of WAI-C to measure change over time and to study differences in alliance ratings between groups

Primary references

1. Busseri MA, Tyler JD. Interchangeability of the working Alliance Inventory and Working Alliance Inventory, Short Form. *Psychol Assess* 2003;**15**:193–7.
2. Cecero JJ, Fenton LR, Frankforter TL, Nich C, Carroll KM. Focus on therapeutic alliance: the psychometric properties of six measures across three treatments. *Psychotherapy* 2001;**38**:1–11.
3. Fenton LR, Cecero JJ, Nich C, Frankforter TL, Carroll KM. Perspective is everything: the predictive validity working alliance instruments. *J Psychother Pract Res* 2001;**10**:262–68.
4. Hanson WE, Curry KT, Bandalos DL. Reliability generalization of Working Alliance Inventory scale scores. *Educ Psychol Measure* 2002;**62**:659–73.
5. Hatcher RL, Barends AW. Patients' view of the alliance in psychotherapy: exploratory factor analysis of three alliance measures. *J Consult Clin Psychol* 1996;**64**:1326–36.
6. Horvath AO, Greenberg LS. The development of the Working Alliance Inventory. In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford; 1986. pp. 529–56.
7. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. *J Counsel Psychol* 1989;**36**:223–33.
8. Safran JD, Wallner LK. The relative predictive validity of two therapeutic alliance measures in cognitive therapy. *Psychol Assess* 1991;**3**:188–95.
9. Stiles WB, Agnew-Davies R, Barkham M, Culverwell A, Goldfried MR, Halstead J, *et al*. Convergent validity of the Agnew Relationship Measure and the Working Alliance Inventory. *Psychol Assess* 2002;**14**:209–20.
10. Tichenor V, Hill CE. A comparison of six measures of working alliance. *Psychotherapy* 1989;**26**:195–99.
11. Tracey TJ, Kokotovic AM. Factor structure of the Working Alliance Inventory. *Psychol Assess* 1989;**1**:207–10.

Secondary reference

12. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychother Theory Res Pract* 1979;**16**:252–60.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

W2
Working Alliance Inventory – Observer (WAI-O)

General details	
Authors	Horvath AO, Greenberg L
Language	English
Country of publication/development	USA
Publication date	NA
Publisher	NA
Purpose and overview	
To assess the strengths and dimensions of the alliance as conceptualised by Bordin[7]	
Theoretical orientation	Pan-theoretical
Population details	Adults
Perspective	Observer
Measure used by	Therapists/counsellors/research clinicians
Other versions	Short Version (12 items) (client/therapist/observer) Working Alliance Inventory – client Working Alliance Inventory – therapist
Notes	
Areas of therapist–patient interaction addressed: Map	
Therapy context: influence; power/coercion; responsibilities	
Roles friend/companion; attachment figure; confidant; expert/authority/leader; protector	
Therapist engagement: all components	
Patient engagement: all components	
Framework: all components	
Therapeutic techniques: responsiveness/receptiveness/attunement; exploration; ruptures/repair	
Threats to the relationship: defensive; critical; fear; resistance; confrontations; withdrawal	
Outcomes: compliance; satisfaction; working alliance; cohesion	
Inferred from scale items	
Dimensions	
Goal agreement	12 items. The extent to which a client and therapist agree on the goals that are the target of the intervention
Task agreement	12 items. The extent to which a client and therapist agree on the in-counselling behaviours and cognitions that form the substance of the counselling relationship
Bond development	12 items. The extent to which a client and therapist possess mutual trust, acceptance and confidence

continued

Reliability	
The WAI-O total scale and dimensions have adequate internal consistency as measured by Cronbach's alpha and adequate inter-rater reliability	
Split-half	NA
Internal consistency	Total scale: 0.98,[4] 0.98,[1] 0.97[3] Dimensions: range 0.93 to 0.97,[1] range 0.84–0.90[3]
Inter-rater	ICC = 0.70,[2] random effects (ICC) = 0.71,[1] inter-rater reliability estimates = 0.79 (range 0.62 to 0.92),[3] ICC = 0.92[4]
Test–retest	No details
Validity	
Extensive validity research has been conducted: the WAI-O has adequate convergent validity and partial predictive validity	
Face	No details
Content	The instrument was generated through a series of sequential ratings and evaluations of prospective items. The initial pool of items was generated on the basis of a content analysis of Bordin's descriptions of the working alliance. The items were rated by experts and professionals. See refs 4 and 5 for details of the item rating summary
Criterion (a) concurrent	No details
Criterion (b) predictive	Correlation between alliance and outcome was 0.39 ($p < 0.001$)[2]
Construct	No details
Convergent	Vanderbilt Therapeutic Alliance Scale (VTAS) $r = 0.87$ ($p < 0.001$), Penn Helping Alliance Scales (Penn) $r = 0.53$ ($p < 0.001$), WAI-C $r = 0.18$ (ns), WAI-T $r = 0.03$ (ns)[2] CALPAS-R $r = 0.82$ ($p < 0.05$), Penn $r = 0.71$ ($p < 0.05$), VTAS $r = 0.75$ ($p < 0.001$), WAI-C $r = 0.21$ (ns), WAI-T $r = 0.03$ (ns)[4] CALPAS-R $r = 0.45$ ($p < 0.001$), Penn $r = 0.50$ ($p < 0.001$), VTAS $r = 0.75$ ($p < 0.001$), WAI-C $r = 0.21$ (ns), WAI-T $r = 0.32$ (ns)[1]
Discriminant	No details
Factor structure	The WAI-O comprises three scales of task, bond and goal. There is substantial overlap among the scales; although the authors have managed to separate the goal and bond scales, the task and goal dimensions share a high degree of covariance. However, the authors found that significant portions of outcome variance that were significantly correlated with the goal or bond scales after the outcome variance explained by the dominant task scale had been removed[5,6] A confirmatory factor analysis[6] found the WAI (Therapist and Client versions) to measure one general alliance factor as well as the three specific alliance dimensions of task, bond and goal. The validity of the three specific aspects of the alliance was more limited than that of the general alliance factor
Responsiveness	
Discriminative (between individuals)	Mean alliance ratings were lowest for clients receiving clinical management treatment for depression[1]
Evaluative (within individual across time)	No details

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Acceptability	
Number of items	36
Administration method	Observer rated
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1986, Guilford, New York
Web or scanning options	No details
Training details	Minimal training
Administration/process details	Assessed after third counseling session in pilot study,[5,6] but may be completed after any counselling/therapy session
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Each of the three subscales has six positive and six negative items rated on a seven-point Likert scale ranging from 1 (never) to 7 (always). Subscale scores range from 12 to 84 and the total score ranges from 36 to 252. Alternatively, a mean score can be taken giving a range of 1 to 7. Higher scores reflect more positive ratings of the working alliance
Normative data	Available in refs 1and 2
Résumé	
Strengths	Adequate convergent validity. High internal consistency and inter-rater reliability. Requires minimal training
Weaknesses	Measure needs a more thorough manual for training purposes
Areas for further research	Use of WAI-O to measure change over time and to study differences in alliance ratings between groups

Primary references

1. Cecero JJ, Fenton LR, Frankforter TL, Nich C, Carroll KM. Focus on therapeutic alliance: the psychometric properties of six measures across three treatments. *Psychotherapy* 2001;**38**:1–11.
2. Fenton LR, Cecero JJ, Nich C, Frankforter TL, Carroll KM. Perspective is everything: the predictive validity working alliance instruments. *J Psychother Pract Res* 2001;**10**:262–8.
3. Hanson WE, Curry KT, Bandalos DL. Reliability generalization of Working Alliance Inventory scale scores. *Educ Psychol Measure* 2002;**62**:659–73.
4. Tichenor V, Hill CE. A comparison of six measures of working alliance. *Psychotherapy* 1989;**26**:195–9.

Secondary references

5. Horvath AO, Greenberg LS. The development of the Working Alliance Inventory. In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford; 1986. pp. 529–56.
6. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. *J Counsel Psychol* 1989;**36**:223–33.
7. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychother Theory Res Pract* 1979;**16**:252–60.

W3
Working Alliance Inventory – Therapist (WAI-T)

General details

Authors	Horvath AO, Greenberg L
Language	English
Country of publication/development	USA
Publication date	1986
Publisher	NA

Purpose and overview

To assess the strengths and dimensions of the alliance as conceptualised by Bordin[10]

Theoretical orientation	Pan-theoretical
Population details	Adults
Perspective	Therapist
Measure used by	Therapists/counsellors/research clinicians
Other versions	Short version (12 items) (available in client, observer and therapist versions) Working Alliance Inventory – Client Working Alliance Inventory – Observer
Notes	

Areas of therapist–patient interaction addressed: Map

Therapy context: influence; power/coercion; responsibilities

Roles: friend/companion; attachment figure; confidant; expert/authority/leader; protector

Therapist engagement: all components

Patient engagement: all components

Framework: all components

Therapeutic techniques: responsiveness/receptiveness/attunement; exploration; ruptures/repair

Threats to the relationship: defensive; critical; fear; resistance; confrontations; withdrawal

Outcomes: compliance; satisfaction; working alliance; cohesion

Information derived from items

Dimensions

Goal agreement	12 items. The extent to which a client and therapist agree on the goals that are the target of the intervention
Task agreement	12 items. The extent to which a client and therapist agree on the in-counselling behaviours and cognitions that form the substance of the counselling relationship
Bond development	12 items. The extent to which a client and therapist possess mutual trust, acceptance and confidence

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
The WAI-T total scale and dimensions have adequate internal consistency as measured by Cronbach's alpha	
Split-half	No details
Internal consistency	Total scale: 0.94 at session 4 and 0.74 at final session,[1] 0.95,[8] 0.95,[2] 0.87,[5,6] 0.91[4] Dimensions: range 0.63 to 0.92,[1] 0.83 to 0.91,[2] 0.68 to 0.87,[5,6] 0.84 to 0.90,[4] 0.90 to 0.93[7]
Inter-rater	Not applicable to client and therapist versions of the WAI
Test–retest	No details
Validity	
Extensive validity research has been carried out on the WAI-T showing the measure to have adequate concurrent and convergent validity and partial predictive validity	
Face	No details
Content	The instrument was generated through a series of sequential ratings and evaluations of prospective items. The initial pool of items was generated on the basis of a content analysis of Bordin's descriptions of the working alliance. The items were rated by experts and professionals. See refs 5 and 6 for details of the item rating summary
Criterion (a) concurrent	For concurrent validity of WAI-T and the WAI-T-S, standardised regression coefficients ranged from 0.88 to 0.99 ($p < 0.01$). After excluding data from clients who unilaterally left therapy, the standardised regression coefficient for the final client (WAI-C) and therapist (WAI-T) ratings was 0.33 ($p < 0.01$)[1]
Criterion (b) predictive	The WAI-T fourth session scores had standardised slope coefficients of 0.40 ($p < 0.01$) for the standardised and averaged 'improvement' score (client and therapist target complaints – improvements and post therapy questionnaire indices) and 0.15 for the standardised and averaged 'symptom' score (client and therapist residual target complaints – severity and client residual global severity index indices)[1] Correlations between alliance and outcome for the WAI-T were not significant[3] The total WAI score was correlated significantly ($r = 0.52$, $p < 0.05$) with the total score of the client post therapy questionnaire (PTQ)[5,6]
Construct	No details
Convergent	CALPAS-R: $r = 0.51$ ($p < 0.001$), Penn Helping Alliance Scales (Penn) $r = 0.44$ ($p < 0.01$), Vanderbilt Therapeutic Alliance Scale (VTAS) $r = 0.36$ (ns), WAI-O $r = 0.36$ (ns), WAI-C $r = 0.43$ ($p < 0.01$)[3] CALPAS-P $r = -0.22$ (ns), Penn $r = 0.20$ (ns), VTAS $r = 0.09$ (ns), WAI-O $r = 0.03$ (ns), WAI-C $r = 0.09$ (ns)[8] CALPAS-R $r = 0.31$ (ns); Penn $r = 0.38$ (ns), VTAS $r = 0.30$ (ns), WAI-O $r = 0.32$ (ns), WAI-C $r = 0.37$ (ns)[2] In a multitrait–multimethod matrix, the WAI scales correlated with the empathy subscale of the Barrett-Lennard Relationship Inventory: range 0.33 to 0.55[5,6]
Discriminant	No details
Factor structure	Random regression coefficients for associations among the subscales ranged from 0.63 to 0.95 ($p < 0.01$)[1] The WAI-C comprises three scales of task, bond and goal. There is substantial overlap among the scales; although the authors have managed to separate the goal and bond scales, the task and goal dimensions share a high degree of covariance. However, the authors found that significant portions of outcome variance that were significantly correlated with the goal or bond scales after the outcome variance explained by the dominant task scale had been removed[5,6]

continued

	A confirmatory factor analysis[9] found the WAI (Therapist and Client versions) to measure one general alliance factor as well as the three specific alliance dimensions of task, bond and goal. The validity of the three specific aspects of the alliance was more limited than that of the general alliance factor
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	The WAI was applied across multiple sessions and session level deviation scores were calculated to provide a measure of session-to-session variation[7]
Acceptability	
Number of items	36
Administration method	Self-report questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1986, Guilford, New York
Web or scanning options	No details
Training details	No details
Administration/process details	Assessed after third counselling session in pilot study,[6,7] but may be completed after any counselling/therapy session
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Each of the three subscales has six positive and six negative items rated on a seven-point Likert scale ranging from 1 (never) to 7 (always). Subscale scores range from 12 to 84 and the total score ranges from 36 to 252. Higher scores reflect more positive ratings of the working alliance
Normative data	Available in ref. 2
Résumé	
Strengths	Adequate convergent validity. High internal consistency and inter-rater reliability. Requires minimal training
Weaknesses	Length of instrument
Areas for further research	Use of WAI-T to measure change over time and to study differences in alliance ratings between groups
Primary references	

1. Busseri MA, Tyler JD. Interchangeability of the working Alliance Inventory and Working Alliance Inventory, Short Form. *Psychol Assess* 2003;**15**:193–7.
2. Cecero JJ, Fenton LR, Frankforter TL, Nich C, Carroll KM. Focus on therapeutic alliance: the psychometric properties of six measures across three treatments. *Psychotherapy* 2001;**38**:1–11.
3. Fenton LR, Cecero JJ, Nich C, Frankforter TL, Carroll KM. Perspective is everything: the predictive validity working alliance instruments. *J Psychother Pract Res* 2001;**10**:262–8.
4. Hanson WE, Curry KT, Bandalos DL. Reliability generalization of Working Alliance Inventory scale scores. *Educ Psychol Measure* 2002;**62**:659–73.

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

5. Horvath AO, Greenberg LS. The development of the Working Alliance Inventory. In Greenberg LS, Pinsof WM, editors. *The psychotherapeutic process: a research handbook. Guilford clinical psychology and psychotherapy series*. New York: Guilford; 1986. pp. 529–56.
6. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. *J Counsel Psychol* 1989;**36**:223–33.
7. Stiles WB, Agnew-Davies R, Barkham M, Culverwell A, Goldfried MR, Halstead J, *et al*. Convergent validity of the Agnew Relationship Measure and the Working Alliance Inventory. *Psychol Assess* 2002;**14**:209–20.
8. Tichenor V, Hill CE. A comparison of six measures of working alliance. *Psychotherapy* 1989;**26**:195–9.
9. Tracey TJ, Kokotovic AM. Factor structure of the Working Alliance Inventory. *Psychol Assess* 1989;**1**:207–10.

Secondary reference

10. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychother Theory Res Pract* 1979;**16**:252–60.

W4
Working Alliance Inventory – Client – Short (WAI-C-S)

General details	
Author	Tracey TJ
Language	English
Country of publication/development	USA
Publication date	1989
Publisher	No details
Purpose and overview	
To assess the strengths and dimensions of the alliance as conceptualised by Bordin.[4] The instrument was primarily designed to sample the therapeutic relationship in its early stages of development (between third and fifth sessions), although applications later in therapy should prove to be equally feasible	
Theoretical orientation	Pan-theoretical,[2] humanistic,[3] psychodynamic,[3] and cognitive-behavioural[3]
Population details	See below
Perspective	Client-rated
Measure used by	Researchers and clinicians
Other versions	Working Alliance Inventory – Therapist – Short Working Alliance Inventory – Observer – Short Working Alliance Inventory – Observer/Client/Therapist (original version)
Notes	*Clients:* The typical study in this review had 56 clients (SD 35). 73% female and 27% male of unknown age. 83% European American and 17% unknown ethnicity, with unknown presenting problems[2] 53 women. Average 22 years[3] *Practitioners:* Psychotherapists,[2] psychologist[3]
Areas of therapist–patient interaction addressed: Map	
Roles: expert/authority/leader	
Therapist engagement: empathy/sensitivity; respect	
Patient engagement: attraction	
Framework: convergent	
Outcomes: achieving a working relationship (task, affective bond, goals, cohesion); general satisfaction	
Inferred directly from items	
Dimensions	
Goal	Four items. Addresses the extent to which therapy goals are important, mutual and capable of being accomplished, e.g. 'The client and therapist have established a good understanding of the changes that would be good for the client'
Task	Four items. Focuses on the participant's agreement about the steps taken to help improve the client's situation, e.g. 'There is agreement on what is important for the client to work on'
Bond	Four items. Measures mutual liking and attachment by focusing on tone of voice, empathy and comfort in exploring intimate issues, e.g. 'There is mutual trust between the client and therapist'

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
Adequate internal consistency has been demonstrated for the WAI-C-S	
Split-half	No details
Internal consistency	At fourth session ratings, Cronbach's alpha was 0.91 (total score), 0.86 (task subscale), 0.73 (goal subscale) and 0.80 (bond subscale). At final session ratings, Cronbach's alpha was 0.92 (total scale), 0.82 (task), 0.81 (goal) and 0.83 (bond)[1] Ranged from 0.92 to 0.98 (mean 0.95, SD 0.03, $n = 3$) for total scores[2] General alliance factor (alpha = 0.98), task factor (alpha = 0.90), bond factor (alpha = 0.92) and goal factor (alpha = 0.90)[3]
Inter-rater	NA
Test–retest	No details
Validity	
Factor analysis demonstrated the bi-level model to have the best fit, although none of the models was found to be a good approximation of the data	
Face	No details
Content	The instrument was generated through a series of sequential ratings and evaluations of prospective items. The initial pool of items was generated on the basis of a content analysis of Bordin's descriptions of the working alliance. Experts and professionals rated the items
Criterion (a) concurrent	For concurrent validity of WAI-C-S and the WAI-C, standardised regression coefficients ranged from 0.80 to 0.97 ($p < 0.01$)[1] After excluding data from clients who unilaterally left therapy, the standardised regression coefficient for the final client (WAI-C-S) and therapist (WAI-T-S) ratings was 0.34 ($p < 0.01$)[1]
Criterion (b) predictive	The WAI-C-S fourth session scores had standardised slope coefficients of 0.34 ($p < 0.01$) for the standardised and averaged 'improvement' score (client and therapist target complaints – improvements and post therapy questionnaire indices) and 0.15 for the standardised and averaged 'symptom' score (client and therapist residual target complaints – severity and client residual global severity index indices)[1]
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	Random regression coefficients for associations among the subscales ranged from 0.73 to 0.92 ($p < 0.01$)[1] Using the factor analysis for the WAI-C, the four highest loading items from each subscale were selected to form a new WAI – Short. The three proposed models of the factor structure (the single general factor – working alliance, the correlated specific factors – goal, task, bond, and the hierarchical belief model – there are three first order factors representing the unique contents as well as a second order, general alliance factor) of the WAI-S were examined using confirmatory factor analysis. The fit criteria for each model was tested in each sample (therapist and client); none of the models was a good approximation of the data. This lack of fit is not surprising given the large number of variables, and the 'fuzzy' nature of the construct. However, the bi-level model had the best relative values, and so is the most appropriate model to represent the data[3]

continued

Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	12
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1989, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	Form completed postsession
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert, seven-point, ranging from 1 (never) to 7 (always). Subscale scores can range from 4 to 28 and can, if desired, be summed to obtain a total score. Thus, total scores can range from 12 to 84. Higher scores reflect more positive ratings of working alliance[2]
Normative data	No details
Notes	
Responses of clients were examined after only one session, which may not be enough time for a working alliance to develop and thus be validly assessed[3]	
Résumé	
Strengths	This short form takes less time to complete than the original WAI-C
	Adequate internal consistency has been demonstrated for the WAI-C-S
Weaknesses	Factor analysis demonstrated the bi-level model to have the best fit, although none of the models was found to be a good approximation of the data
Areas for further research	Future cross-validation work is needed to support the current results[3]
	Further study of psychometric properties is recommended

Primary references

1. Busseri MA, Tyler JD. Interchangeability of the working Alliance Inventory and Working Alliance Inventory, Short Form. *Psychol Assess* 2003;**15**:193–7.
2. Hanson WE, Curry KT, Bandalos DL. Reliability generalization of Working Alliance Inventory scale scores. *Educ Psychol Measure* 2002;**62**:659–73.
3. Tracey TJ, Kokotovic AM. Factor structure of the Working Alliance Inventory. *Psychol Assess* 1989;**1**:207–10.

Secondary reference

4. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychother Theory Res Pract* 1979;**16**:252–60.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

W5
Working Alliance Inventory – Observer – Short (WAI-O-S)

General details

Author	Tracey TJ
Language	English
Country of publication/development	USA
Publication date	1989
Publisher	NA

Purpose and overview

To assess the strengths and dimensions of the alliance as conceptualised by Bordin.[2] The instrument was primarily designed to sample the therapeutic relationship in its early stages of development (between third and fifth sessions), although applications later in therapy should prove to be equally feasible

Theoretical orientation	Designed to cover a range of therapies
Population details	Clinical adults. See below
Perspective	Observer-rated
Measure used by	Researchers and clinicians
Other versions	Working Alliance Inventory – Client – Short Working Alliance Inventory – Therapist – Short Working Alliance Inventory – Observer/Client/Therapist (original version)
Notes	*Clients:* Average age 39 years. Female to male ratio 3.5:1. More than 80% Caucasian. Met criteria for major depression in DSM, scored ⩾20 on BDI, and scored ⩾14 on Hamilton Rating Scale for Depression[1] *Practitioners:* Psychotherapist[1] *Raters:* One psychology student (who had 25 hours of training) and one psychology graduate (with extensive rating experience)[1]

Areas of therapist–patient interaction addressed: Map

Roles: expert/authority/leader

Therapist engagement: empathy/sensitivity; respect

Patient engagement: attraction

Framework: convergent

Outcomes: achieving a working relationship (task, affective bond, goals, cohesion); general satisfaction

Inferred directly from items

Dimensions

Goal	Four items. Addresses the extent to which therapy goals are important, mutual and capable of being accomplished, e.g. 'The client and therapist have established a good understanding of the changes that would be good for the client'
Task	Four items. Focuses on the participant's agreement about the steps taken to help improve the client's situation, e.g. 'There is agreement on what is important for the client to work on'
Bond	Four items. Measures mutual liking and attachment by focusing on tone of voice, empathy and comfort in exploring intimate issues, e.g. 'There is mutual trust between the client and therapist'

continued

Reliability	
Adequate internal consistency was reported. Overall score inter-rater reliability was partially adequate. Item-by-item inter-rater reliabilities ranged from inadequate to partial	
Split-half	No details
Internal consistency	0.81, as reported in Gelfand and DeRubeis (unpublished manuscript)[1]
Inter-rater	0.67. Item-by-item inter-rater reliabilities ranged from 0.14 to 0.65[1]
Test–retest	No details
Validity	
A two-factor structure of the WAI-O-S is supported, with factor 1 (agreement/confidence) accounting for 58.4% of the variance and factor 2 (relationship) accounting for 15% of the variance	
Face	No details
Content	The instrument was generated through a series of sequential ratings and evaluations of prospective items. The initial pool of items was generated on the basis of a content analysis of Bordin's descriptions of the working alliance. Experts and professionals rated the items
Criterion (a) concurrent	No details
Criterion (b) predictive	No details
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	A principal components analysis revealed two independent factors. The principal components eigenvalues were 7 and 1.8, accounting for 73.4% of the variance (58.4% and 15%, respectively). Factor 1 was labelled 'agreement/confidence' and consists of four goal items, four task items and one bond item. Factor 2 was labelled 'relationship' and consists of the remaining three bond items[1]
Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	12 items: ten positively worded, two negatively worded
Administration method	Observer-rated questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1989, American Psychological Association
Web or scanning options	No details
Training details	Training was completed by using sample sessions of CBT that the two raters rated separately, comparing ratings afterwards. Training proceeded until the two raters had a similar understanding of the scale items and scoring procedures, and until the reliability between the two raters was deemed acceptable[1]

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Administration/process details	The WAI-O-S was completed for each of the 70 tapes of session 2 by each of the two raters. Ratings were made independently after listening to an entire session of therapy then averaged, and the raters were blind to the identity of the patient and therapist and to the eventual outcomes of each case[1]
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, seven-point Likert scale
Normative data	No details
Résumé	
Strengths	This short form takes less time to complete than the original WAI-O
	Adequate internal consistency was reported. Overall score inter-rater reliability was partially adequate
	A two-factor structure of the WAI-O-S is supported
Weaknesses	With only 12 items, the authors may have missed a more precise conceptualisation of the construct of alliance in CBT
	Item-by-item inter-rater reliabilities ranged from inadequate to partially adequate
Areas for further research	Further investigation of psychometric properties is recommended
Primary reference	
1. Andrusyna TP, Tang TZ, DeRubeis RJ, Luborsky L. The factor structure of the Working Alliance Inventory in cognitive-behavioral therapy. *J Psychother Pract Res* 2001;**10**:173–8.	
Secondary reference	
2. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychother Theory Res Pract* 1979;**16**:252–60.	

W6
Working Alliance Inventory – Therapist – Short (WAI-T-S)

General details	
Author	Tracey TJ
Language	English
Country of publication/development	USA
Publication date	1989
Publisher	No details
Purpose and overview	
To assess the strengths and dimensions of the alliance as conceptualised by Bordin.[4] The instrument was primarily designed to sample the therapeutic relationship in its early stages of development (between third and fifth sessions), although applications later in therapy should prove to be equally feasible	
Theoretical orientation	Psychodynamic,[3] humanistic,[3] cognitive behavioural[3] and various/range[2]
Population details	Clinical adults. See notes
Perspective	Therapist-rated
Measure used by	Researchers and practitioners
Other versions	Working Alliance Inventory – Client – Short Working Alliance Inventory – Observer – Short Working Alliance Inventory – Observer/Client/Therapist (original version)
Notes	*Clients:* The typical study had 56 clients (SD 35). 73% female and 27% male of unknown age. 83% European American and 17% unknown ethnicity, with unknown presenting problems[2] 53 women. Average 22 years[3] *Practitioners:* Psychologist,[3] psychotherapist[2]
Areas of therapist–patient interaction addressed: Map	
Roles: expert/authority/leader	
Therapist engagement: empathy/sensitivity; respect	
Patient engagement: attraction	
Framework: convergent	
Outcomes: achieving a working relationship (task, affective bond, goals, cohesion); general satisfaction	
Inferred directly from items	
Dimensions	
Goal	Four items. Addresses the extent to which therapy goals are important, mutual and capable of being accomplished, e.g. 'The client and therapist have established a good understanding of the changes that would be good for the client'
Task	Four items. Focuses on the participant's agreement about the steps taken to help improve the client's situation, e.g. 'There is agreement on what is important for the client to work on'
Bond	Four items. Measures mutual liking and attachment by focusing on tone of voice, empathy, and comfort in exploring intimate issues, e.g. 'There is mutual trust between the client and therapist'

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Reliability	
Adequate internal consistency has been demonstrated for the WAI-T-S	
Split-half	No details
Internal consistency	At fourth session ratings, Cronbach's alpha was 0.91 (total score), 0.89 (task subscale), 0.81 (goal subscale), and 0.77 (bond subscale). At final session ratings, Cronbach's alpha was 0.96 (total scale), 0.90 (task), 0.90 (goal) and 0.86 (bond)[1]
	Ranged from 0.90 to 0.95 (mean 0.93, SD 0.04, $n = 2$) for total scores[2]
	General alliance factor (alpha = 0.95), task factor (alpha = 0.83), bond factor (alpha = 0.91) and goal factor (alpha = 0.88)[3]
Inter-rater	No details
Test–retest	No details
Validity	
Factor analysis demonstrated the bi-level model to have the best fit, although none of the models was found to be a good approximation of the data	
Face	No details
Content	The instrument was generated through a series of sequential ratings and evaluations of prospective items. The initial pool of items was generated on the basis of a content analysis of Bordin's descriptions of the working alliance. Experts and professionals rated the items
Criterion (a) concurrent	For concurrent validity of WAI-T-S and the WAI-T, standardised regression coefficients ranged from 0.88 to 0.99 ($p < 0.01$)[1]
	After excluding data from clients who unilaterally left therapy, the standardised regression coefficient for the final client (WAI-C-S) and therapist (WAI-T-S) ratings was 0.34 ($p < 0.01$)[1]
Criterion (b) predictive	The WAI-T fourth session scores had standardised slope coefficients of 0.40 ($p < 0.01$) for the standardised and averaged 'improvement' score (client and therapist target complaints – improvements and post therapy questionnaire indices) and 0.15 for the standardised and averaged 'symptom' score (client and therapist residual target complaints – severity and client residual global severity index indices)[1]
Construct	No details
Convergent	No details
Discriminant	No details
Factor structure	Random regression coefficients for associations among the subscales ranged from 0.63 to 0.95 ($p < 0.01$)[1]
	Using the factor analysis for the WAI-C, the four highest loading items from each subscale were selected to form a new WAI – Short. The three proposed models of the factor structure (the single general factor – working alliance, the correlated specific factors – goal, task, bond, and the hierarchical bi-level model – there are three first order factors representing the unique contents as well as a second order, general alliance factor) of the WAI-S were examined using confirmatory factor analysis. The fit criteria for each model was tested in each sample (therapist and client); none of the models was a good approximation of the data. This lack of fit is not surprising given the large number of variables, and the 'fuzzy' nature of the construct. However, the bi-level model had the best relative values, and so is the most appropriate model to represent the data[3]

continued

Responsiveness	
Discriminative (between individuals)	No details
Evaluative (within individual across time)	No details
Acceptability	
Number of items	12
Administration method	Questionnaire
Time taken to complete	No details
Flesch reading age	No details
Translations	No details
Access by ethnic minorities	No details
Feasibility	
Copyright	1989, American Psychological Association
Web or scanning options	No details
Training details	No details
Administration/process details	Form completed postsession
Support from measure developers	No details
FAQ facility	No details
Precision	
Scale type	Ordinal, Likert, seven-point, ranging from 1 (never) to 7 (always). Subscale scores can range from 4 to 28 and can, if desired, be summed to obtain a total score. Thus, total scores can range from 12 to 84. Higher scores reflect more positive ratings of working alliance[2]
Normative data	No details
Notes	
Responses of clients were examined after only one session, which may not be enough time for a working alliance to develop and thus be validly assessed[3]	
Résumé	
Strengths	This short form takes less time to complete than the original WAI-T
	Adequate internal consistency has been demonstrated for the WAI-T-S
Weaknesses	Factor analysis demonstrated the bi-level model to have the best fit although none of the models was found to be a good approximation of the data
Areas for further research	Future cross-validation work is needed to support the current results[3]
	Further investigation of psychometric properties is recommended

Primary references

1. Busseri MA, Tyler JD. Interchangeability of the Working Alliance Inventory and Working Alliance Inventory, Short Form. *Psychol Assess* 2003;**15**:193–7.
2. Hanson WE, Curry KT, Bandalos DL. Reliability generalization of Working Alliance Inventory scale scores. *Educ Psychol Measure* 2002;**62**:659–73.
3. Tracey TJ, Kokotovic AM. Factor structure of the Working Alliance Inventory. *Psychol Assess* 1989;**1**:207–10.

Secondary reference

4. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychother Theory Res Pract* 1979;**16**:252–60.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Appendix 9

Content description of candidate measures

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Measure	ID	Author, year	Areas of TPI addressed	Theoretical orientation	Perspective	Population groups	No. of items	Dimensions
Affective Sensitivity Scale – Form A	A1	Campbell, 1971	TE, OC	Interpersonal	Therapist	Therapists	86	0
Affective Sensitivity Scale – Form C	A2	Campbell, 1971	TE, OC	Interpersonal	Therapist	Therapists	89	0
Affective Sensitivity Scale – Form D	A3	Kagan, 1987	TE	Interpersonal	Therapist	Therapists	No details	0
Affective Sensitivity Scale – Form D-80	A4	Kagan, 1987	TE	Interpersonal	Therapist	Therapists	63	6
Affective Sensitivity Scale – Forms E-80 and E-A-2	A5	Kagan, 1987	TE	Interpersonal	Therapist	Therapists	57	6
Affective Sensitivity Scale – Form H	A6	Kagan, 1994	TE	Interpersonal	Therapist	Therapists	No details	1
Agnew Relationship Measure	A7	Agnew-Davies, 1998	TC, R, TE, PE, F, TT, TTR, OC	Pan-theoretical	Therapist, patient	Adults	28	4
Barrett-Lennard Relationship Inventory	B1	Barrett-Lennard, 1962	TE, F, TT, NVC, OC, AWR, EE	Person-centred/ pan-theoretical	Patient, therapist, observer	Adults, groups, children, adolescents	64	4
California Psychotherapy Alliance Scale – Original	C1	Marmar, 1989	TC, R, TE, PE, F, TT, TTR, OC	Pan-theoretical	Patient, therapist, observer	Adults	Pt 31, Th 5, Ob 5	5
California Psychotherapy Alliance Scales – Patient	C2	Marmar, 1991	TC, R, TE, PE, F, TT, TTR, OC	Pan-theoretical	Patient	Adults	24	4
California Psychotherapy Alliance Scales – Rater	C3	Marmar, 1991	TC, R, TE, PE, F, TT, TTR, OC	Pan-theoretical	Observer	Adults	24	4
California Psychotherapy Alliance Scales – Therapist	C4	Marmar, 1991	TC, R, TE, PE, F, TT, TTR, OC	Pan-theoretical	Therapist	Adults	24	4
California Therapeutic Alliance Rating System	C5	Marmar, 1981	TC, TE, PE, F, TT	Psychoanalytic/ pan-theoretical	Observer	Adults	41	4
California Therapeutic Alliance Rating System Scales	C6	Marmar, 1984	TC, TE, PE, F, TT	Psychoanalytic/ pan-theoretical	Observer, patient, therapist	Adults	42	4
Capacity for Dynamic Process Scale	C7	Baumann, 2001	FW, EE, AWR, CSO	Psychodynamic	Therapist, observer	Adults	9	0
Carkhuff Scales	C8	Carkhuff, 1969	TE, TTR	Person-centred	Observer	Adults	NA	NA
Child Psychotherapy Process Scales	C9	Estrada, 1996	TE, PE, TT, AWR	Psychodynamic	Observer	Children	33	6
Client Attachment to Therapist Scale	C10	Mallinckrodt, 1995	ID, PE	Psychoanalytic	Patient	Adults	36	3

continued

Measure	ID	Author, year	Areas of TPI addressed	Theoretical orientation	Perspective	Population groups	No. of items	Dimensions
Client Resistance Scale	C11	Mahalik, 1994	PE, TTR, EE	Psychoanalytic/ pan-theoretical	Observer	Adults	5	5
Coding the Interaction in Psychotherapy	C12	Schindler, 1989	TE, PE, F, TT, NVC, TTR, OC	Behavioural	Observer	Adults	37	13
Coherence of the Relationship Theme	C13	Mitchell, 1995	PE	Psychodynamic	Observer	Adults	NA	3
Core Conflictual Relationship Theme	C14	Luborsky, 1976	ID, PE, TT, OC, CSO	Psychoanalytic	Observer	Adults	NA	3
Counseling Evaluation Inventory	C15	Linden, 1965	TC, R, TE, F, TT, TTR, OC	Counselling	Patient	Adults	21	3
Counsellor Effectiveness Rating Scale	C16	Atkinson, 1982	R, PE, TTR, AWR, GS	Counselling	Observer	Therapists	10	4
Counsellor Effectiveness Scale	C17	Ivey, 1971	TC, TE, R, TT, TTR, OC, F	Counselling/ pan-theoretical	Observer	Therapists	25	3
Counsellor Rating Form	C18	Barak, 1975	TC, R, TE, PE	Counselling	Patient, observer	Adults, adolescents	36	3
Counsellor Rating Form Short Version	C19	Corrigan, 1983	R, TE, PE	Counselling	Patient, observer	Adults, adolescents	12	3
Counselor Evaluation Rating Scale	C20	Myrick, 1971	TE, F, NVC, TTR, OC	Counselling/ pan-theoretical	Therapist, observer	Therapists	27	3
Counselor Perception Questionnaire	C21	Blocher, 1985	NVC	Counselling	Observer	Adults	NA	2
Cross-Cultural Counseling Inventory	C22	La Frombroise, 1991	BC, TC, R, ID, TE, F, OC	Counselling	Observer	Therapists	20	3
Empathy Construct Rating Scale – 23	E1	Hughes, 1982	TC, TE, F, TT, TTR	Mental health nursing	Patient, therapist, peer observer	Adults	23	0
Empathy Construct Rating Scale – 84	E2	La Monica, 1981	TC, TE, F, TT, TTR	Mental health nursing	Patient, therapist, peer observer	Adults	84	2
Empathy test	E3	Layton, 1979	TE, TT	Person-centred	Therapist	Therapists	17	0
Experiencing Scale	E4	Clein, 1979	PE, TTR, OC	Pan-theoretical	Observer	Adults	1	0

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Measure	ID	Author, year	Areas of TPI addressed	Theoretical orientation	Perspective	Population groups	No. of items	Dimensions
Family Engagement Questionnaire	F1	Kroll, 1997	TC, R, ID, TE, PE, F, TTR	Psychiatry	Therapist	Families	16	4
Family Therapeutic Alliance Scale	F2	Martin, 1993	NVC, TE, PE, F, TT, TTR, AWR	Pan-theoretical	Observer	Families	15	2
Feminist Self-Disclosure Inventory	F3	Simi, 1997	TC, R, TE, F, TT, TTR, OC	Feminism	Therapist	Therapists	20	5
Group Assessment of Interpersonal Traits	G1	Goodman, 1972	TE, F, TT	Interpersonal	Patient, observer, peer patient	Groups	7	7
Helper Behaviour Rating System (Modified)	H1	Shapiro, 1984	TE, TT	Person-centred/ pan-theoretical	Observer	Adults	12	0
Helpful Responses Questionnaire	H2	Miller, 1991	TE, TT	Not specified	Therapist	Therapists	6	0
Helping Alliance Counting Signs Method	H3	Luborsky, 1983	TC, ID, TE, PE, F, OC	Psychoanalytic	Observer	Adults	NA	7
Hill Client Verbal Response Category System	H4	Hill, 1981	ID, F, PE, NVC, O	Pan-theoretical	Observer	Adults	9	0
Hill Interaction Matrix Form G	H5	Hill, 1975	R, F, PE, TTR, OC	Pan-theoretical	Observer, therapist, patient	Groups	72	10
Hill Interaction Matrix Statement by Statement	H6	Hill, 1965	R, F, PE, TTR, OC	Pan-theoretical	Observer	Adults, groups, children	20	9
Hill Verbal Counselor Response Category System	H7	Hill, 1978	TE, F, TT, NVC, TTR	Pan-theoretical	Observer	Therapists	14	0
Hill Verbal Counselor Response Category System – Revised	H8	Friedlander, 1982	R, TE, TT, F, TTR	Pan-theoretical	Observer	Therapists	9	0
Integrative Psychotherapy Alliance Scale	I1	Pinsof, 1986	R, ID, TE, PE, F, TT, AWR, WA	Systemic	Patient	Families	25	2
Inter-Session Experience Scale	I2	Orlinsky, 1993	TC, R, ID, PE, F, TTR, OC	Psychodynamic	Patient	Adults	42	3
Maslach Burnout Inventory – Client and Therapist versions	M1	Linehan, 2000	TC, TE, PE, F, TT, TTR, AWR	Not specified	Patient, therapist	Adults	22	3

continued

Measure	ID	Author, year	Areas of TPI addressed	Theoretical orientation	Perspective	Population groups	No. of items	Dimensions
Missouri Identifying Transference Scale	M2	Multon, 1996	TC, R, ID, PE, TT, TTR, OC	Psychoanalytic/ psychodynamic	Therapist	Adults	43	2
Multicultural Counseling Inventory	M3	Sodowsky, 1994	TC, R, TE	Counselling	Therapist	Adults	40	4
Octant Scale Impact Message Inventory	O1	Keisler, 1997	TC, R, PE, F, TTR	Interpersonal	Therapist, patient	Adults	56	8/2
Patient Action Scale	P1	Hoyt, 1981	TC, R, ID, PE, F, NVC	Psychodynamic	Observer	Adults	24	0
Penn Helping Alliance Questionnaire	P2	Woody, 1983	R, PE, F, GS, AWR	Pan-theoretical	Patient, observer	Adults	11	2
Penn Helping Alliance Questionnaire – Revised	P3	Luborsky, 1996	R, F, TE, PE, TTR, OC	Pan-theoretical	Patient, therapist	Adults	19	0
Penn Helping Alliance Rating Scale	P4	Luborsky, 1986	TE, F, GS, AWR, CSO	Pan-theoretical	Observer	Adults	10	2
Psychotherapy Process Inventory	P5	Baer, 1980	TC, R, ID, PE, F, TT, TTR, GS	Pan-theoretical	Therapist	Adults	74	4
Psychotherapy Process Q-Set	P6	Jones, 2000	TC, R, ID, TE, PE, F, TT, NVC, TTR, OC	Psychodynamic/ pan-theoretical	Observer	Adults	100	0
Reasons for Ending Treatment Questionnaire	R1	Garcia, 2002	TC, R, ID, TE, PE, TTR, OC	Not specified	Therapist, parent	Children, adolescents	41	6
Session Evaluation Questionnaire	S1	Stiles, 1980	OC	Psychodynamic	Patient, therapist, observer	Adults	22	3
Session Evaluation Questionnaire Form 3	S2	Stiles, 1984	F, OC	Psychodynamic	Patient, therapist	Adults	24	4
Session Evaluation Questionnaire Form 4	S3	Stiles, 1994	R, TE, F, OC	Psychodynamic	Patient, therapist	Adults	27	5
Session Impacts Scale	S4	Elliott, 1994	TC, R, ID, TE, PE, F, TT, TTR, AWR, EE, CSO	Process experiential	Patient	Adults	17	3
Therapeutic Alliance Scales for Children	T1	Shirk, 1992	F, OC	Pan-theoretical	Patient, therapist	Children	Pt 8, Th 7	3

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Measure	ID	Author, year	Areas of TPI addressed	Theoretical orientation	Perspective	Population groups	No. of items	Dimensions
Therapeutic Bond Scales	T2	Saunders, 1989	TE, PE, AWR, EE	Psychodynamic	Patient	Adults	50	3
Therapeutic Factors Inventory	T3	Lese, 2000	TC, R, PE, F, TT, TTR, OC	Interpersonal	Patient	Groups	99	11
Therapist Action Scale	T4	Hoyt, 1981	TC, R, ID, TE, PE, F, NVC	Psychodynamic	Therapist, observer	Adults	25	0
Therapist Behavior Scale	T5	Duckro, 1980	TC, R, TE, PE, FW	Not specified	Patient	Adults	40	2
Therapist Representation Inventory – 4th Section: Record of Dreams	T6	Geller, 1982	All areas applicable	Psychodynamic	Patient	Therapists	NA	NA
Therapist Representation Inventory – Free Response Task	T7	Geller, 1982	All areas applicable	Psychodynamic	Patient	Therapists	NA	NA
Therapist Representation Inventory – Therapist Embodiment Scale	T8	Geller, 1982	TC, R, ID, TE, TT, NVC	Psychodynamic	Patient	Therapists	12	3
Therapist Representation Inventory – Therapist Involvement Scale	T9	Geller, 1982	TC, R, ID, TE, PE, F, NVC, TTR, OC	Psychodynamic	Patient	Therapists	38	6
Truax and Carkhuff Scales	T10	Truax, 1967	R, TE, F, TT, TTR	Person-centred	Observer	Adults	NA	NA
Vanderbilt Negative Indicators Scale	V1	Strupp, 1986	PE, TE, TTR	Psychodynamic	Therapist	Adults	42	3
Vanderbilt Negative Indicators Scale – Short	V2	Nergaard, 1989	PE, TE, TTR	Psychodynamic	Observer	Adults	No details	3
Vanderbilt Psychotherapy Process Scale – 80 item	V3	Strupp, 1983	TC, R, ID, TE, PE, F, TT, NVC, TTR, EE, CSO	Pan-theoretical	Observer	Adults	80	8
Vanderbilt Therapeutic Alliance Scale	V4	Hartley, 1983	TC, R, TE, PE, F, TT, NVC, TTR, CSO, GS, AWR	Pan-theoretical	Observer	Adults	44	6
Working Alliance Inventory – Client	W1	Horvath, 1986	TC, R, TE, PE, F, TT, TTR, OC	Pan-theoretical	Patient	Adults	36	3
Working Alliance Inventory – Observer	W2	Horvath, 1986	TC, R, TE, PE, F, TT, TTR, OC	Pan-theoretical	Observer	Adults	36	3
Working Alliance Inventory – Therapist	W3	Horvath, 1986	TC, R, TE, PE, F, TT, TTR, OC	Pan-theoretical	Therapist	Adults	36	3

continued

Measure	ID	Author, year	Areas of TPI addressed	Theoretical orientation	Perspective	Population groups	No. of items	Dimensions
Working Alliance Inventory: Client Short Form	W4	Tracey, 1989	R, TE, PE, F, AWR, GS	Pan-theoretical	Patient	Adults	12	3
Working Alliance Inventory: Observer Short Form	W5	Andrusyna, 1989	R, TE, PE, F, AWR, GS	Pan-theoretical	Observer	Adults	12	3
Working Alliance Inventory: Therapist Short Form	W6	Tracey, 1989	R, TE, PE, F, AWR, GS	Pan-theoretical	Therapist	Adults	12	3

AWR, achieving a working relationship; BC, broader context; CSO, changing view of self with others; EE, emotional expression; F, framework; GS, general satisfaction; ID, individual differences; NA, not applicable; NVC, non-verbal communication; OC, outcomes; PE, patient engagement; R, roles; TT, therapeutic techniques; TTR, threats to relationship; TC, therapy context; TE, therapist engagement; TPI, therapist–patient interaction.
Therapists: focus is on the practitioner; adults: individuals >18 years attending therapy; groups/families: patients in group/family therapy; children/adolescents: individuals <18 years attending therapy; adult inpatients: individuals >18 years in hospital psychiatric settings.

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Appendix 10

Psychometric properties of candidate measures

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Measure	ID	Reliability	Validity	Responsiveness	Acceptability	Feasibility	Precision
Affective Sensitivity Scale – Form A	A1	SH = Partial IC = Partial	Conv = Partial	Not addressed	Partially addressed 2	Partially addressed 2	Partially addressed
Affective Sensitivity Scale – Form C	A2	SH = Partial TR = Adequate	Const = Adequate Conv = Partial[a] Pred = Partial[a]	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Affective Sensitivity Scale – Form D	A3	IC = Adequate TR = Partial	Conv = Partial	Adequate	Partially addressed 1	Partially addressed 2	Partially addressed
Affective Sensitivity Scale – Form D-80	A4	IC = Partial	Face = Addressed Conv = Adequate	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Affective Sensitivity Scale – Forms E-80 and E-A-2	A5	TR = Partial[a] IC = Partial	Face = Addressed Conv = Partial[a]	Not addressed	Partially addressed 2	Partially addressed 2	Partially addressed
Affective Sensitivity Scale – Form H	A6	IC = Adequate TR = Adequate	Pred = Adequate	Adequate	Partially addressed 2	Partially addressed 4	Partially addressed
Agnew Relationship Measure	A7	IC = Adequate	Face = Addressed Cont = Addressed Pred = Partial Conv = Adequate	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Barrett-Lennard Relationship Inventory	B1	IC = Adequate TR = Adequate Observer: IR = Inadequate	Face = Addressed Cont = Adequate Pred = Adequate Cons = Adequate Conv = Partial[a] FS = Addressed	Adequate	Partially addressed 4	Partially addressed 2	Partially addressed
California Psychotherapy Alliance Scale – Original	C1	IC = Adequate IR = Adequate	Pred = Partial Const = Adequate Conv = Partial FS = Addressed	Adequate	Partially addressed 2	Partially addressed 3	Partially addressed
California Psychotherapy Alliance Scales – Patient	C2	IC = Adequate	Conc = Partial Pred = Partial Const = Adequate Conv = Partial Disc = Adequate FS = Addressed	Partial	Partially Addressed 4	Partially Addressed 4	Partially Addressed

continued

Measure	ID	Reliability	Validity	Responsiveness	Acceptability	Feasibility	Precision
California Psychotherapy Alliance Scales – Rater	C3	IC = Adequate IR = Adequate	Conv = Adequate Cont = Adequate Pred = Adequate Conc = Adequate Disc = Adequate FS = Addressed	Adequate	Partially addressed 3	Partially addressed 4	Addressed
California Psychotherapy Alliance Scales – Therapist	C4	No details	Face = Addressed Conv = Adequate FS = Addressed	Not addressed	Partially addressed 3	Partially addressed 4	Partially addressed
California Therapeutic Alliance Rating System	C5	IC = Adequate IR = Partial[a]	Face = Addressed Cont = Addressed Conv = Partial Conc = Partial Pred = Partial[a] Disc = Adequate FS = Addressed	Adequate	Partially addressed 3	Partially addressed 3	Partially addressed
California Therapeutic Alliance Rating System Scales	C6	IC = Adequate IR = Partial	Face = Addressed Pred = Partial Cons = Adequate Conv = Partial FS = Addressed	Adequate	Partially addressed 3	Partially addressed 3	Partially addressed
Capacity for Dynamic Process Scale	C7	IC = Adequate IR = Adequate	Conv = Partial Disc = Adequate	Not addressed	Partially addressed 2	Partially addressed 2	Partially addressed
Carkhuff Scales	C8	IR = Adequate	Face = Addressed Conc = Adequate Pred = Partial Conv = Partial Cons = Inadequate	Adequate	Partially addressed 2	Partially addressed 3	Partially addressed
Child Psychotherapy Process Scales	C9	IC = Adequate IR = Adequate	FS = Addressed	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Client Attachment to Therapist Scale	C10	IC = Partial TR = Adequate	Face = Addressed Cont = Adequate Conc = Partial FS = Addressed	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Client Resistance Scale	C11	IR = Partial	Cons = Adequate Conv = Adequate FS = Addressed	Adequate	Partially addressed 2	Partially addressed 3	Partially addressed

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Measure	ID	Reliability	Validity	Responsiveness	Acceptability	Feasibility	Precision
Coding the Interaction in Psychotherapy	C12	IR = Adequate	Conv = Partial Cont = Adequate	Adequate	Partially addressed 3	Partially addressed 3	Partially addressed
Coherence of the Relationship Theme	C13	IR = Partial	Conv = Partial	Adequate	Partially addressed 1	Partially addressed 3	NA
Core Conflictual Relationship Theme	C14	IR = Partial	Face = Addressed Conc = Partial Pred = Partial Cons = Adequate Conv = Adequate	Adequate	Partially addressed 3	Partially addressed 4	addressed
Counseling Evaluation Inventory	C15	IC = Adequate TR = Adequate	Cont = Adequate Conc = Partial Cons = Adequate Conv = Partial FS = Addressed	Adequate	Partially addressed 4	Partially addressed 3	Partially addressed
Counselor Effectiveness Rating Scale	C16	IC = Adequate	Conc = Adequate Pred = Adequate Conv = Adequate FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 1	Adequate
Counsellor Effectiveness Scale	C17	IR = Inadequate	Cont = Addressed FS = Addressed	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Counselor Rating Form	C18	IC = Adequate SH = Adequate IR = Adequate	Face = Addressed Cont = Addressed Pred = Adequate Cons = Partial Conv = Partial[a] FS = Addressed	Partial[a]	Partially addressed 4	Partially addressed 2	Partially addressed
Counselor Rating Form – Short Version	C19	SH = Adequate IC = Adequate	Face = Addressed Cont = Addressed Conc = Partial[a] Disc = Inadequate FS = Addressed	Adequate	Partially addressed 5	Partially addressed 2	Partially addressed
Counselor Evaluation Rating Scale	C20	SH = Adequate IC = Adequate TR = Adequate	Face = Addressed Cont = Addressed Conv = Partial FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 4	Partially addressed

continued

Measure	ID	Reliability	Validity	Responsiveness	Acceptability	Feasibility	Precision
Counselor Perception Questionnaire	C21	IR = Adequate	Cont = Addressed Conv = Adequate Disc = Adequate	Partial	Partially addressed 3	Partially addressed 2	Partially addressed
Cross-Cultural Counseling Inventory	C22	IC = Adequate IR = Partial[a]	Face = Addressed Cont = Addressed Disc = Adequate Cons = Adequate FS = Addressed	Adequate	Partially addressed 4	Partially addressed 2	Partially addressed
Empathy Construct Rating Scale – 23	E1	IC = Adequate	Conv = Partial Cons = Adequate	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Empathy Construct Rating Scale – 84	E2	SH = Adequate IC = Adequate	Cont = Adequate Conc = Inadequate Disc = Adequate FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 2	Partially addressed
Empathy Test	E3	SH = Partial	Face = Addressed Cont = Addressed Conv = Partial	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Experiencing Scale	E4	IR = Adequate	Face = Addressed Conc = Partial Pred = Partial Cons = Partial Conv = Partial Disc = Adequate	Adequate	Partially addressed 3	Partially addressed 3	Addressed
Experiencing Scale	E4	IR = Adequate	Face = Addressed Conc = Partial Pred = Partial Cons = Partial Conv = Partial Disc = Adequate	Adequate	Partially addressed 3	Partially addressed 3	Addressed
Family Engagement Questionnaire	F1	IC = Partial IR = Partial	Face = Addressed Cont = Addressed Conv = Partial FS = Addressed	Adequate	Partially addressed 2	Partially addressed 3	Partially addressed
Family Therapeutic Alliance Scale	F2	IC = Partial TR = Adequate IR = Adequate	Face = Addressed Cont = Addressed FS = Addressed	Adequate	Partially addressed 3	Partially addressed 2	Partially addressed

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Measure	ID	Reliability	Validity	Responsiveness	Acceptability	Feasibility	Precision
Feminist Self-Disclosure Inventory	F3	IC = Adequate TR = Adequate	Cont = Adequate Cons = Adequate FS = Addressed	Adequate	Partially addressed 2	Partially addressed 3	Partially addressed
Group Assessment of Interpersonal Traits	G1	SH = Partial IC = Partial IR = Partial TR = Adequate	Pred = Partial Cons = Adequate Conv = Partial	Adequate	Partially addressed 2	Partially addressed 4	Partially addressed
Helper Behaviour Rating System (modified)	H1	IR = Adequate	Cont = Addressed	Not addressed	Partially addressed 2	Partially addressed 3	Partially addressed
Helpful Responses Questionnaire	H2	IC = Partial IR = Adequate TR = Inadequate	Conv = Inadequate	Adequate	Partially addressed 3	Partially addressed 3	Addressed
Helping Alliance Counting Signs Method	H3	IR = Partial[a]	Cont = Addressed Conc = Adequate Pred = Adequate Conv = Partial[a]	Adequate	Partially addressed 3	Partially addressed 4	Partially Addressed
Hill Client Verbal Response Category System	H4	IR = Adequate	Face = Addressed Cont = Addressed	Partial	Partially addressed 3	Partially addressed 4	Addressed
Hill Interaction Matrix – Form G	H5	IR = Partial[a] IC = Partial	Face = Addressed Cont = Addressed Conv = Adequate FS = Addressed	Not addressed	Partially addressed 3	Partially addressed 3	Addressed
Hill Interaction Matrix – Statement by Statement	H6	IR = Adequate	Face = Addressed Cont = Addressed Conv = Adequate	Adequate	Partially addressed 2	Partially addressed 4	Addressed
Hill Counselor Verbal Response Category System	H7	IR = Partial[a]	Face = Addressed Cont = Adequate Conv = Partial	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Hill Counselor Verbal Response Category System – Revised	H8	IR = Partial[a]	Face = Addressed Cont = Adequate Conv = Partial	Adequate	Partially addressed 2	Partially addressed 3	Partially addressed
Integrative Psychotherapy Alliance Scale	I1	IC = Adequate TR = Adequate	Pred = Adequate	Partial	Partially addressed 3	Partially addressed 4	Partially addressed
Intersession Experience Questionnaire	I2	IC = Partial	Face = Addressed FS = Addressed	Adequate	Partially addressed 2	Partially addressed 1	Partially addressed

continued

Measure	ID	Reliability	Validity	Responsiveness	Acceptability	Feasibility	Precision
Maslach Burnout Inventory – Client and Therapist versions	M1	IC = Partial	Face = Addressed Cont = Addressed Pred = Partial Conv = Partial Disc = Adequate FS = Addressed	Adequate	Partially addressed 3	Partially addressed 3	Addressed
Missouri Identifying Transference Scale	M2	IC = Adequate	Conc = Partial FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 3	Partially addressed
Multicultural Counseling Inventory	M3	IC = Adequate	Face = Addressed Cont = Adequate Conv = Adequate FS = Addressed	Adequate	Partially addressed 4	Partially addressed 2	Partially addressed
Octant Scale Impact Message Inventory	O1	IC = Partial	Conv = Partial FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 2	Partially addressed
Patient Action Scale	P1	IR = Partial TR = Partial	Face = Addressed Cont = Addressed Conv = Partial FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 3	Partially addressed
Penn Helping Alliance Questionnaire	P2	No details	Cont = Addressed Pred = Adequate Conv = Adequate FS = Addressed	Adequate	Partially addressed 2	Partially addressed 5	Addressed
Penn Helping Alliance Questionnaire – Revised	P3	Patient: IC = Adequate TR = Adequate Therapist: IC = Adequate TR = Partial	Conv = Adequate Disc = Partial	Inadequate	Partially addressed 2	Partially addressed 2	Partially addressed
Penn Helping Alliance Rating Scale	P4	IC = Partial[a] TR = Partial IR = Partial[a]	Pred = Partial[a] Cons = Adequate Conv = Partial[a] Disc = Partial[a]	Partial	Partially addressed 2	Partially addressed 4	Addressed
Psychotherapy Process Inventory	P5	IC = Adequate	Face = Addressed Cont = Adequate Pred = Partial FS = Addressed	Adequate	Partially addressed 3	Partially addressed 2	Partially addressed

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Measure	ID	Reliability	Validity	Responsiveness	Acceptability	Feasibility	Precision
Psychotherapy Process Q-Set	P6	IC = Adequate IR = Adequate	Face = Addressed Cont = Adequate Pred = Partial Conv = Inadequate FS = Addressed	Adequate	Partially addressed 2	Partially addressed 4	Addressed
Reasons for Ending Treatment Questionnaire	R1	IC = Adequate TR = Adequate	Pred = Partial FS = Addressed	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Session Evaluation Questionnaire	S1	IC = Adequate	Face = Addressed Client: Pred = Inadequate Observer and therapist: Pred = Partial FS = Addressed	Adequate	Partially addressed 2	Partially addressed 2	Partially Addressed
Session Evaluation Questionnaire – Form 3	S2	IC = Adequate	Face = Addressed Cont = Addressed Conv = Partial Disc = Inadequate FS = Addressed	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Session Evaluation Questionnaire – Form 4	S3	IC = Adequate	Face = Addressed Cont = Addressed Client: Pred = Partial FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 2	Partially addressed
Session Impacts Scale	S4	IC = Partial[a]	Face = Addressed Conc = Partial Cons = Adequate Disc = Adequate FS = Addressed	Partial	Partially addressed 2	Partially addressed 2	Partially addressed
Therapeutic Alliance Scales for Children	T1	IC = Partial	Cont = Addressed Conv = Partial FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 2	Partially addressed
Therapeutic Bond Scales	T2	IC = Partial	Cont = Addressed Pred = Adequate	Not addressed	Partially addressed 3	Partially addressed 4	Partially addressed

continued

Measure	ID	Reliability	Validity	Responsiveness	Acceptability	Feasibility	Precision
Therapeutic Factors Inventory	T3	IC = Adequate TR = Partial	Face = Addressed Cont = Adequate Cons = Adequate Conv = Partial FS = Addressed	Adequate	Partially addressed 2	Partially addressed 2	Partially addressed
Therapist Action Scale	T4	Two raters: IR = Adequate Rater and therapist: IR = Partial	Face = Addressed Cont = Addressed Cons = Adequate Conv = Adequate FS = Addressed	Not addressed	Partially addressed 3	Partially addressed 3	Partially addressed
Therapist Behavior Scale	T5	TR = Adequate	Cont = Addressed FS = Addressed	Partial	Partially addressed 2	Partially addressed 1	Partially addressed
Therapist Representation Inventory – 4th Section: Record of Dreams	T6	–	–	–	–	–	–
Therapist Representation Inventory – Free Response Task	T7	–	–	–	–	–	–
Therapist Representation Inventory – Therapist Embodiment Scale	T8	IC = Partial	Pred = Inadequate FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 2	Partially addressed
Therapist Representation Inventory – Therapist Involvement Scale	T9	IC = Adequate	Pred = Partial FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 2	Partially addressed
Truax and Carkhuff (1967) Scales	T10	2/3 scales: IR = Adequate	Conv = Partial[a]	Partial	Partially addressed 2	Partially addressed 4	Partially addressed
Vanderbilt Negative Indicators Scale	V1	IC = Partial IR = Partial	Pred = Adequate	Not addressed	Partially addressed 2	Partially addressed 1	Partially addressed
Vanderbilt Negative Indicators Scale – Short	V2	IC = Partial IR = Adequate	Pred = Partial Conv = Partial	Not addressed	Partially addressed 1	Partially addressed 1	Partially addressed
Vanderbilt Psychotherapy Process Scale – 80 item	V3	IR = Adequate IC = Adequate	Pred = Partial FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 4	Partially addressed
Vanderbilt Therapeutic Alliance Scale	V4	IC = Adequate IR = Partial[a]	Conc = Partial Pred = Partial Conv = Partial	Adequate	Partially addressed 2	Partially addressed 3	Partially addressed

continued

© Queen's Printer and Controller of HMSO 2008. All rights reserved.

Measure	ID	Reliability	Validity	Responsiveness	Acceptability	Feasibility	Precision
Working Alliance Inventory – Client	W1	IC = Adequate TR = Adequate	Cont = Adequate Conc = Adequate Conv = Adequate Pred = Partial[a] Disc = Adequate FS = Addressed	Adequate	Partially addressed 2	Partially addressed 2	Addressed
Working Alliance Inventory – Observer	W2	IC = Adequate IR = Adequate	Cont = Adequate Pred = Partial Conv = Adequate FS = Addressed	Adequate	Partially addressed 2	Partially addressed 3	Addressed
Working Alliance Inventory – Therapist	W3	IC = Adequate	Cont = Adequate Pred = Partial[a] Conv = Partial FS = Addressed	Adequate	Partially addressed 2	Partially addressed 2	Addressed
Working Alliance Inventory – Client – Short	W4	IC = Adequate	FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 1	Partially addressed
Working Alliance Inventory – Observer – Short	W5	IC = Adequate IR = Partial	FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 2	Partially addressed
Working Alliance Inventory – Therapist – Short	W6	IC = Adequate	FS = Addressed	Not addressed	Partially addressed 2	Partially addressed 1	Partially addressed

Reliability and validity judgements are based on total scores where possible.
[a] 'Partial' validity is a function of the variability in findings across multiple studies.
Conc, Concurrent; Cons, Construct; Const, constant; Conv, convergent; Disc, discriminant; Face, face validity; FS, factor structure; IC, internal consistency; IR, inter-rater; Pred, predictive; SH, split-half; TR, test–retest.

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.hta.ac.uk) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

The National Coordinating Centre for Health Technology Assessment,
Alpha House, Enterprise Road
Southampton Science Park
Chilworth
Southampton SO16 7NS, UK.
Fax: +44 (0) 23 8059 5639   Email: hta@hta.ac.uk
http://www.hta.ac.uk

ISSN 1366-5278