

Measurement Criteria for Review of Free Measures

An objective of this project was to identify strong, practical measures that are appropriate for initial assessment and progress monitoring of internalizing, externalizing, and/or substance abuse symptoms. The following measures met most or all of these criteria:

1. Reliable*
2. Valid**
3. Brief (fewer than 50 items)
4. Free or low cost
5. Sensitive to change (for progress and outcomes monitoring)
6. Used in schools
7. Validated for use with urban, low-SES, minority youth

***Reliability:** We categorized reliability data for each measure (internal consistency and test-retest reliability) as Adequate, Good, or Excellent based on the criteria set forth by Hunsley & Mash (2008):

	Research evidence indicates:
<i>Internal Consistency</i>	
Adequate	α values of .70-.79
Good	α values of .80-.89
Excellent	α values of \geq .90
<i>Test-retest Reliability</i>	
Adequate	Test-retest correlations of at least .70 over a period of several days to several weeks.
Good	Test-retest correlations of at least .70 over a period of several months.
Excellent	Test-retest correlations of at least .70 over a period of a year or longer.

****Validity:** We categorized validity data for each measure based on the following criteria developed with our population in mind (urban, minority, low-income youth in the school setting):

High: Demonstrates concurrent, predictive, convergent, and divergent validity. Predictive of academic performance or failure. Validated with low-income, urban, minority youth.

Medium: Most types of validity demonstrated or measured, however, no research on predictive validity for academic performance and/or validity with low-income, urban, minority youth.

Low: Few or no types of validity demonstrated or measured. No research on predictive validity for academic performance and/or validity with low-income, urban, minority youth

Matrix of Selected Mental Health and Substance Use Measures for School Mental Health

Instrument* ***	Purpose***				Domains Measured	Age (yrs)	Forms**			# of Items	Time (mins)	Reliability*		Validity
	S	I	P	T			P	Y	T			T-R	IC	
BPC	✓	✓	✓	✓	Internalizing, externalizing problems	7-18	✓	✓		12-15	5	Adeq.	Adeq.	Medium
CANS-MH	✓	✓	✓	✓	Problem presentation, risk behaviors, functioning, child safety, caregiver needs & strengths, and child/adolescent strengths	0-18				47	5-15	---	---	Medium
CRAFFT	✓	✓			Alcohol, drug use	14-21		✓		3-9	1-2	Poor	Adeq.	Medium
GAIN-SS	✓	✓	✓		Internalizing externalizing, substance use, crime/violence problems	12-adult		✓		23	3-5	---	Exc.	Medium
POSIT	✓	✓	✓	✓	Substance abuse, physical & mental health, family & peer relations, educational & vocational status, social skills, leisure, & aggression/delinquency	12-19		✓		139	20-30	Adeq.	Adeq. to Good	Medium

PSC/ PSC-17	✓	✓	✓	✓	Attention, internalizing, externalizing problems	P: 4-16 Y:11-18	✓	✓	17-35	5	Adeq.	Exc.	High	
SDQ	✓	✓	✓	✓	Emotional symptoms, conduct problems, hyperactivity, inattention, peer problems & prosocial behavior	P: 3-16 Y:11-17 T: 3-16	✓	✓	✓	25	5-10	Poor	Adeq.	Medium
SEI	✓	✓	✓		Affective & cognitive factors related to school engagement	12-18		✓		35	5	Poor	Adeq. to Exc.	High
YTP			✓	✓	Primary concerns to be addressed in treatment	7-13	✓	✓		3	5-8	Adeq.	---	Medium

****Full names of each instrument on following page

***S=Screening; IA=Initial Assessment; PM=Treatment Monitoring; TO=Treatment Outcome

**P=Parent, Y=Youth, T=Teacher

*T-R=Test-Retest; IC=Internal Consistency; Adeq.=Adequate; Exec.=Excellent

Detailed Description of Selected Mental Health and Substance Use Measures (1 pagers)

1. Brief Problem Checklist (BPC; Chorpita, Reise, Weisz, Grubbs, Becker, & Krull, 2010)
2. Child and Adolescent Needs and Strengths: An Information Integration Tool for Children and Adolescents with Mental Health Challenges (CANS-MH; Lyons, Griffin, Fazio, & Lyons, 1999)
3. CRAFFT (Knight, Shrier, Bravender, Farrell, Vanderbilt, & Shaffer, 1999)
4. The Global Appraisal of Individual Needs-Short Screener (GAIN-SS; Dennis, Chan, & Funk, 2006)
5. The Problem Oriented Screening Instrument for Teenagers (POSIT; Rahdert, 1991)
6. Pediatric Symptom Checklist (PSC & Y-PSC; Jellinek, Murphy, Robinson, Feins, Lamb, & Fenton, 1988; PSC-17; Gardner, Murphy, Childs, Kelleher, Pagano, Jellinek, et al., 1999)
7. Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997)
8. Student Engagement Instrument (SEI; Appleton, Christenson, Kim, & Reschly, 2006)
9. Youth Top Problems (YTP; Weisz, Chorpita, Frye, Ng, Lau, Bearman, et al., 2011)

(Note: On the 1-pagers, ^G = This term can be found in the Glossary of Terms)

Brief Problem Checklist (BPC)

Description

The Brief Problem Checklist (BPC) is a questionnaire that measures externalizing and internalizing problems. It is designed for repeated periodic assessment of clinical progress and was originally designed to be administered over the phone. The BPC was developed from the Youth Self-Report and Child Behavior Checklist (YSR and CBCL, Achenbach & Rescorla, 2001) so it represents a free of cost, much shortened variation of these tools.

Purpose

Screening Initial Assessment Progress Monitoring Treatment outcome

Sample Items

- Argues a lot Unhappy, sad, or depressed
- Disobedient at home or at school Worries

Response Options

Not True, Somewhat True, Very True

Time frame

In the last week

Forms

Parent	Youth	Teacher
Ages 7-18	Ages 7-18	N/A
12-15 Items	12-15 Items	

Completion Time

5 minutes

Statistical Strength ^G

Reliability ^G: Adequate (test-retest ^G, internal consistency ^G).

Validity ^G: Medium (concurrent ^G, predictive ^G, convergent ^G, divergent ^G).

Research with: Schools, urban youth, minority youth, low-SES youth.

Standardization: Cut-off scores ^G Based on normative samples ^G

Languages Available

English

Where to Access

Child version: <http://www.childfirst.ucla.edu/Brief%20Problem%20Checklist%20-%20Child.pdf>

Parent Version: <http://www.childfirst.ucla.edu/Brief%20Problem%20Checklist%20-%20Parent.pdf>

Certification/Cost

None/Free

Developers

Bruce Chorpita, Ph.D., John Weisz, Ph.D., Steven Reise, Ph.D.

Child and Adolescent Needs and Strengths

An Information Tool for Children and Adolescents with Mental Health Challenges (CANS-MH)

Description

The CANS-MH is a measure completed by the clinician following an initial intake interview or case review. The clinician provides ratings across six domains: Problem Presentation, Risk Behaviors, Functioning, Child Safety, Caregiver Needs & Strengths, and Child/Adolescent Strengths to directly inform treatment planning, intensity, and monitoring.

Purpose

Screening Initial Assessment Progress Monitoring Treatment outcome

Sample Items

- Attention Deficit/Impulse
- Substance Abuse
- Danger to Self
- Family strengths

Response Options

No evidence, Mild degree, Moderate degree, Severe or profound degree, Unknown

Time frame

Past 30 days

Forms

Parent	Youth	Teacher
N/A*	N/A*	N/A*

*The CANS-MH (47 items; Ages 0-18) is clinician-administered.

Completion Time

5-15 minutes

Statistical Strength ^G

Reliability ^G: Good inter-rater reliability. Test-retest ^G and internal consistency ^G reliability unknown.

Validity ^G: Medium (concurrent ^G, predictive ^G, convergent ^G).

Research with: Urban youth, minority youth, low-SES youth.

Standardization: Cut-off scores ^G Based on normative samples ^G

Languages Available

English, Spanish, Portugese

Where to Access

<http://www.magellanprovider.com/education/outcomes-library/child-and-adolescent-needs-and-strengths-cans-mh/cans-training.aspx>
CANStraining.com

Certification/Cost

Only Praed Foundation can provide certification. \$10/person for annual subscription for training and certification.

Developers

John S. Lyons, Ph.D.

CRAFFT (Car, Relax, Alone, Forget, Friends, Trouble)

Description

CRAFFT is a questionnaire that assesses youth alcohol and other drug use. Meeting the cut-off score indicates need for more assessment and a longer conversation about the context of use, frequency, and other risks and consequences of alcohol and other drug use is warranted.

Purpose

Screening Initial Assessment Progress Monitoring Treatment outcome

Sample Items

- | | |
|---|--|
| <p>C - Have you ever ridden in a CAR driven by someone (including yourself) who was "high" or had been using alcohol or drugs?</p> <p>R - Do you ever use alcohol or drugs to RELAX, feel better about yourself, or fit in?</p> <p>A - Do you ever use alcohol/drugs while you are by yourself, ALONE?</p> | <p>F - Do you ever FORGET things you did while using alcohol or drugs?</p> <p>F - Do your family or FRIENDS ever tell you that you should cut down on your drinking or drug use?</p> <p>T - Have you gotten into TROUBLE while you were using alcohol or drugs?</p> |
|---|--|

Response Options Yes or No

Time frame In the past year (screener); "Have you ever?" (CRAFFT questions)

Forms

Parent	Youth	Teacher
N/A	Ages 14-21 3 screener items, then 6 items	N/A

Completion Time 1-2 minutes

Statistical Strength ^G

Reliability ^G: Poor test-retest ^G; adequate internal consistency ^G.

Validity ^G: Medium (high sensitivity and specificity for detecting presence of a problem, disorder, or dependence).

Research with: Schools, urban youth, minority youth, low-SES youth.

Standardization: Cut-off scores ^G Based on normative samples ^G

Languages Available 13 languages (e.g., English, Spanish, Chinese, French)

Where to Access www.ceaser-boston.org/CRAFFT/

Cost Free
Developers Children's Hospital Boston

Global Appraisal of Individual Needs, Short Screener (GAIN-SS)

Description

Global Appraisal of Individual Needs, Short Screener (GAIN-SS) is a questionnaire used to screen for behavioral health disorders (e.g., internalizing or externalizing psychiatric disorders, substance use disorders, and crime/violence problems). It is a shortened version of the 123-item GAIN, and shows high agreement with the GAIN.

Purpose

Screening Initial Assessment Progress Monitoring Treatment outcome

Sample Items

- When was the last time that you had significant problems with thinking about ending your life or committing suicide?
- When was the last time that you did the following things two or more times? (e.g., Had a hard time waiting your turn)
- When was the last time that you used alcohol or other drugs weekly or more often?
- When was the last time that you took something from a store without paying for it?

Response Options

Past month, 2-3 months ago, 4-12 months, 1+ years ago, Never

Time frame

N/A

Forms

Parent	Youth	Teacher
N/A	Ages 12-Adult 23 items	N/A

Completion Time

3-5 minutes

Statistical Strength ^G

Reliability ^G: No test-retest reported; excellent internal consistency ^G.

Validity ^G: Medium (predictive ^G, divergent ^G, high sensitivity and specificity for detecting presence of a disorder).

Research with: Schools, urban youth, minority youth, low-SES youth.

Standardization: Cut-off scores ^G Based on normative samples ^G

Languages Available

English, Spanish

Where to Access

<http://www.gaincc.org/>

Certification/Cost Developers

Licensing \$100 per agency; covers 5 years of paper assessments.
Chestnut Health Systems

^G= This term can be found in the Glossary of Terms

Problem Oriented Screening Instrument for Teenagers (POSIT)

Description

The Problem Oriented Screening Instrument for Teenagers (POSIT) is a questionnaire used to identify potential problems in 10 areas: substance use and abuse, physical health, mental health, family relations, peer relations, educational status, vocational status, social skills, leisure/recreation, and aggressive behavior/delinquency. The POSIT was designed to be administered by a variety of assessors, including school personnel.

Purpose

Screening Initial Assessment Progress Monitoring Treatment outcome

Sample Items

- Do you brag? Do you threaten to hurt people?
- Do you often feel tired? Are you a good reader?

Response Options

Yes or No

Time Frame

Not specified

Forms

Parent	Youth	Teacher
N/A	Ages 12-19	N/A
	POSIT (139 items)	
	POSIT-Follow up (89 items)	

Completion Time

20-30 minutes

Statistical Strength ^G

Reliability ^G: Adequate test-retest ^G; adequate-good internal consistency ^G.

Validity ^G: Medium (concurrent ^G, predictive ^G, convergent ^G, divergent ^G; high sensitivity and specificity in identifying diagnostic classification).

Research with: Schools, urban youth, minority youth, low-SES youth.

Standardization: Cut-off scores ^G Based on normative samples ^G

Languages Available

English, Spanish

Where to Access

<http://www.emcdda.europa.eu/html.cfm/index4439EN.html>

Certification/Cost

None/Free

Developers

Elizabeth Rahdert, Ph.D.; National Institute on Drug Abuse,
National Institutes of Health

^G= This term can be found in the Glossary of Terms

Pediatric Symptom Checklist (PSC)

Description

The Pediatric Symptom Checklist (PSC) is a psychosocial questionnaire that consists of three subscales: Attention Problems, Internalizing Problems, and Externalizing Problems. While it was initially developed for use in health care settings, its application has been expanded to school-based health services.

Purpose

Screening Initial Assessment Progress Monitoring Treatment outcome

Sample Items

- Seems to be having less fun Fights with others
- Fidgety, unable to sit still Worries a lot

Response Options Never, Sometimes, or Often

Time frame None specified

Forms

Parent (Ages 4-16)	Youth (Ages 11-18)	Teacher
Ages 4-16	Ages 11-18	N/A
PSC (35 items)	PSC-Y (35 items)	
PSC-17 (17 items)	PSC-17-Y (17 items)*	

* Total score and subscale scores not yet validated

Completion Time 5 minutes

Statistical Strength ^G

Reliability ^G: Adequate test-retest ^G; excellent internal consistency ^G.

Validity ^G: High (concurrent ^G, predictive ^G, convergent ^G, divergent ^G; high specificity^G and sensitivity ^G when compared to clinicians' ratings of children's psychosocial dysfunction; predicts academic failure; validated for use with urban, low-SES, African-American youth).

Research with: Schools, urban youth, minority youth, low-SES youth.

Standardization: Cut-off scores ^G Based on normative samples ^G

Languages Available 18 languages (e.g., English, Spanish, French, Chinese)

Where to Access http://www.massgeneral.org/psychiatry/services/psc_home.aspx

Certification/Cost None/Free

Developers Michael Jellinek, M.D.; Michael Murphy, Ed.D.

^G= This term can be found in the Glossary of Terms

Strengths and Difficulties Questionnaire (SDQ)

Description

The Strength and Difficulties Questionnaire (SDQ) is a behavioral questionnaire that screens for emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior. In one study, it was identified as the most practical quantitative instrument for use in schools because it is multi-informant and can be administered to a wide age range (Borntager & Lyon, 2014)

Purpose

Screening Initial Assessment Progress Monitoring Treatment outcome

Sample Items

- Often unhappy, depressed or tearful
- Restless, overactive, cannot stay still for long
- Often fights with other children or bullies other
- Many fears, easily scared

Response Options

Not True, Somewhat True, or Certainly True

Time frame

In the last 6 months

Forms

Parent	Youth	Teacher
Ages 3-16	Ages 11-17	Ages 3-16
SDQ (25 items)	SDQ (25 items)	SDQ (25 items)
Impact Supplement (5 items)*	Impact Supplement (5 items)*	Impact Supplement (5 items)*
Follow-up Version (2 items)**	Follow-up Version (2 items)**	Follow-up Version (2 items)**

* Assesses problem chronicity, distress, social impairment, and burden to others.

** Assess symptoms improvement and perceived helpfulness of treatment.

Completion Time

5-10 minutes

Statistical Strength ^G

Reliability ^G: Poor test-retest ^G; adequate internal consistency ^G.

Validity ^G: Medium (concurrent ^G, predictive ^G, convergent ^G).

Research with: Schools, urban youth, minority youth, low-SES youth.

Standardization: Cut-off scores ^G Based on normative samples ^G

Languages Available

82 languages (e.g., English, Spanish, French, Chinese)

Where to Access

<http://www.sdqinfo.org/>

Certification/Cost

None/Free

Developers

Robert Goodman, M.D.

^G= This term can be found in the Glossary of Terms

Student Engagement Instrument (SEI)

Description

The Student Engagement Instrument (SEI) is a questionnaire used to measure a student's level of engagement at school and with learning. The measure examines affective engagement factors (i.e., teacher-student relationships, peer support for learning, family support for learning) and cognitive engagement factors (i.e., control and relevance of school work, future goals and aspirations, extrinsic motivation).

Purpose

Screening Initial Assessment Progress Monitoring Treatment outcome

Sample Items (*Strongly Disagree, Disagree, Agree, and Strongly Disagree*)

- Other students at school care about me
- The tests in my classes do a good job of measuring what I'm able to do
- Going to school after high school is important
- I'll learn, but only if the teacher gives me a reward

Response Options Yes or No

Time frame None Specified

Forms

Parent	Youth	Teacher
N/A	Grades 6-12	N/A
	35 items	

Completion Time 5 minutes

Statistical Strength ^G

Reliability ^G: Poor test-retest ^G; adequate to excellent internal consistency ^G.

Validity ^G: High (concurrent ^G, predictive ^G, convergent ^G, divergent ^G). Correlated with GPA, behavioral incidents, and student achievement. Predictive of dropout and performance on the state mandated high-stakes achievement assessment.

Research with: Schools, urban youth, minority youth, low-SES youth.

Standardization: Cut-off scores ^G Based on normative samples ^G

Languages Available English

Where to Access <http://checkandconnect.umn.edu/research/engagement.html>

Certification/Cost None/Free

Developers James Appleton, Ph.D., Sandra Christenson, Ph.D., Dongjin Kim, Amy Reschly, Ph.D.

^G= This term can be found in the Glossary of Terms

Youth Top Problems (YTP)

Description

Youth Top Problems (YTP) is a structured way of assessing client and/or parent report of primary concerns to be addressed in treatment. It can be used flexibly to monitor treatment progress and inform decisions about when to end treatment.

Purpose

Screening Initial Assessment Progress Monitoring Treatment outcome

Sample Items

- List the problems you are most concerned about
 - How big of a problem is this for you on a scale ranging from 0 (*not at all*) to 10 (*very, very much*)
- Which one is the biggest problem right now? Which of these is giving you the most trouble right now? Which one is the most important to work on? (go through each identified problem until all ranked)

Response Options N/A

Time frame “Right now”

Forms

Parent	Youth	Teacher
Ages 7-13	Ages 7-13	N/A
YTP (3 items)	YTP (3 items)	

Completion Time 5-8 minutes

Statistical Strength ^G

Reliability ^G: Adequate test-retest ^G.

Validity ^G: Medium (concurrent ^G, predictive ^G, convergent ^G, divergent ^G).

Research with: Lack of research.

Standardization: Cut-off scores ^G Based on normative samples ^G

Languages Available English

Where to Access

[http://www.childfirst.ucla.edu/Weisz%20et%20al%20\(2011\)-Youth%20Top%20Problems.pdf](http://www.childfirst.ucla.edu/Weisz%20et%20al%20(2011)-Youth%20Top%20Problems.pdf) (Note: There is no template form available)

Certification/Cost None/Free

Developers John Weisz, Ph.D., Bruce Chorpita, Ph.D., Alice Frye, Ph.D., et al.

^G= This term can be found in the Glossary of Terms

Glossary of Psychometric/ Measurement Terms

Concurrent validity: association of a measure with a benchmark test, like comparing a new IQ test to an already established and validated IQ test.

Construct validity: defines how well a test measures the construct it claims to measure. A test designed to measure depression must only measure that particular construct, not closely related constructs such as anxiety or stress.

Convergent validity: the degree to which a scale is associated with measures of similar constructs even when they are measured with a different modality. A test designed to measure depression should be correlated with other measures of depression.

Criterion validity: association of a test measure with a criterion variable (other measures or outcomes already considered valid). For example, IQ tests are often validated by examining their correlation with measures of academic performance.

Cut-off scores: a reference point, usually numerical, used to divide scores into two or more classifications, with some action to be taken (e.g., further evaluation necessary) or some inference to be made (e.g., no concern, at risk, clinically significant).

Divergent validity: the degree to which a scale is not associated with measures of different constructs. A test designed to measure depression must only measure that particular construct, not closely related constructs such as anxiety or stress.

Internal consistency: a measure of reliability used to evaluate the degree to which different test items on the measure that probe the same construct produce similar results.

Normative sample: a group of people presumed to be representative of the people who may be assessed by a measure. This group's scores on the measure may be used as a reference source for evaluating how an individual's scores compares to other's scores (for example, of the same grade, age, ethnicity, etc.).

Predictive validity: a measure's ability to predict something it should theoretically be able to predict. For example, an IQ should be able to predict scores of academic performance.

Reliability: the degree to which an assessment yields stable or consistent results. There are different types of reliability such as the stability of a measure when administered twice to the same group of people (test-retest reliability), how much different respondents agree (e.g., interrater reliability among parent, student, and clinician reports), whether results from different versions are similar (e.g., parallel forms reliability) and/or whether items within the assessment

^G= This term can be found in the Glossary of Terms

"hang" together or are strongly related (e.g., internal consistency reliability often expressed as Cronbach's alpha).

Sensitivity (a.k.a, true positive rate): Measures the percentage of individuals that are correctly identified as having the condition (e.g., symptom, disorder, ect.) by the measure.

Specificity (a.k.a, true negative rate): Measures the percentage of individuals that are correctly identified as NOT having the condition (e.g., symptom, disorder, ect.) by the measure.

Statistical strength: refers to the psychometric properties of a particular tool or measure, based on the research published about the measure. Statistical strength includes reliability, validity, and other important findings that tell us about the quality of the tool based on research.

Test-retest reliability: a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals. The scores from Time 1 and Time 2 are then correlated to evaluate the test for stability over time.

Validity: refers to the degree to which an assessment measures what it is "supposed" to or designed to measure. In other words, does the assessment accurately capture the construct or topic of interest? Types of validity include when stakeholders can "tell" what the assessment is trying to assess based on the items (e.g., face validity), when the measure captures what it is "supposed" to and doesn't capture other constructs (e.g., construct validity), when the measure can predict current or future performance (e.g., criterion-related validity), when the measure is able to help improve the program of study (e.g., formative validity), and when the measure ensures the full content area is adequately captured (e.g., sampling validity).