Evidence-Based Assessment Tools for Common Mental Health Problems: A Practical Guide for School Settings

Jack H. Andrews, Evelyn Cho, Siena K. Tugendrajch, Brigid R. Marriott, & Kristin M. Hawley

Department of Psychological Sciences, University of Missouri – Columbia

Author Note

The authors have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Jack H. Andrews, Department of Psychological Sciences, 3 McAlester Hall, 320 S 6th St., Columbia, MO 65211. Email: andrewsjh@missouri.edu

The Version of Record of this manuscript has been published and is available in Children and Schools Published by Oxford University Press: https://doi.org/10.1093/cs/cdz024

Abstract

Evidence-based assessment (EBA), which requires the use of reliable and valid measurement tools, is an essential component of many services that school social workers provide to promote the social, emotional, and behavioral health of students. A wide variety of psychometrically-sound assessment tools exist to choose from, but it can be difficult for school personnel to identify and access the tools best suited to meet their needs. In an effort to reduce these barriers, the present article provides a concise guide to free, validated measurement tools for the most common youth mental health concerns (anxiety, depression, and disruptive behavior) that are feasible for routine use in school settings. The psychometric properties and other practical characteristics of thirty-seven measurement tools are reviewed, accompanied by links to access each tool and suggestions to help social workers identify which may best fit any particular combination of the diverse goals, school settings, and student populations they serve.

Keywords: evidence-based assessment, school mental health, measurement tools

Evidence-Based Assessment Tools for Common Mental Health Problems: A Practical Guide for School Settings

Promoting the social, emotional, and behavioral health of students is a primary focus of school social workers and other school-based mental health professionals. Evidence-based assessment (EBA), which requires the use of reliable and valid measurement tools, is a foundational component of quality care and effective mental health services, from prevention to intervention (APA Task Force on Evidence-Based Practice with Children and Adolescents, 2008). EBA is essential to achieve optimal outcomes for the student, classroom, school and community. EBA may be useful for school- or district-wide surveillance of student mental health concerns, to identify at-risk students in need of prevention or early intervention, to screen for problems that require more intensive services, guide clinical decision making throughout intervention, evaluate intervention effectiveness, and demonstrate strengths and needs to diverse stakeholders including regulatory bodies, funding agencies, students, families, and taxpayers. Furthermore, evidence-based monitoring and assessment may also deliver therapeutic benefits in its own right (e.g., Tam & Ronan, 2017; Thompson, 2014).

Given the immense diversity of the populations and settings served by school social workers, the myriad ways that students may be affected by mental health concerns, and the multiple functions EBA must serve for schools, it is fortunate that numerous psychometrically-sound, youth mental health assessment tools are available. Unfortunately, schools can provide unique challenges to routine implementation of EBA. Notable barriers include difficulty selecting, scoring, or interpreting measures; financial costs to purchase measures; time and money required to train school personnel to administer a measure; lost educational time required

for students to complete a measure; and poor fit between specific measurement tools and the target population, setting, or function (Connors, Arora, Curtis, & Stephan, 2015).

Highlighting the need for EBA to be practical, Beidas and colleagues (2015) conducted a review of standardized assessment instruments that were not just validated, but also free and brief. Our aim is to update and expand upon their work by providing a concise guide to EBA tools for school social workers and other school mental health personnel. This guide includes (a) a brief overview of the different properties of measures that may be most important to consider when selecting an EBA tool, (b) tips for efficiently investigating different measures and choosing between them, and (c) a review of the practical characteristics of specific, psychometrically-sound measurement tools that are free and feasible to use for school-based assessment of one or more of the most common youth mental health concerns: anxiety, depression, and disruptive behavior (Merikangas et al., 2010a and 2010b).

We developed this guide to help school social workers quickly identify tools to assess common mental health concerns and begin using these tools in practice. For some, a refresher on psychometrics may be worthwhile (e.g., Brooks & Kutcher, 2001; Furr, 2017; Myers & Winters, 2002). Given space limitations and the immense volume of the research literature, we do not present all information school social workers might wish to consider for their specific setting or population (i.e., we do not summarize all available evidence for each potential application of the reviewed tools). Moreover, any attempt to do so would almost immediately become outdated (e.g., psychometric properties of existing measures are tested in new populations, new measures are developed, existing measures are revised). Instead, our review provides a snapshot to compare and contrast current tools for assessing common mental health concerns, as well as links and guidance to help readers quickly search for additional detail beyond that presented

here. In addition, our overview of measure characteristics and tips for evaluating and selecting specific tools are broadly applicable to EBA in any domain relevant to mental health (e.g., other mental health concerns; strengths and competencies). We also provide links and citations to other published reviews covering additional tools and details that may be relevant to school mental health practice but fell outside the scope of the current review.

How to Use This Article to Choose a Measurement Tool

To use this guide to select a measurement tool, first determine your purpose. Carefully consider how you will use the measure and note any critical or desirable features (e.g., youth self-report; takes under 5 minutes to complete; sensitive to change). Below we review specific characteristics that may be particularly important to consider. Next, identify a 'shortlist' of tools that have these characteristics. Table 1 provides an overview of tools and their specific characteristics. To ensure that you have the most up-to-date information, we also recommend reviewing the author's or the measure's website whenever available. The authors of many of the assessment tools maintain detailed websites full of information and resources related to the measure (e.g., copies of the measures themselves; scoring manuals; some even provide free, downloadable scoring templates). When this proves insufficient, or appears out-of-date, you may also search for published research that cites the measure (e.g., via Google Scholar). Article abstracts are free to view and often provide important psychometric information. In the sections that follow, we provide recommendations for specific search terms and strategies to efficiently search for information about each of a tool's specific characteristics. Finally, compare the specific characteristics of each measurement tool on your shortlist and review the measures themselves in order to determine which is best suited to your needs.

Specific Characteristics of Measurement Tools to Evaluate

We divide characteristics of measurement tools into two broad categories: psychometric properties (i.e., the evidence base supporting its reliability and validity) and other practical characteristics that may influence a tool's suitability for a specific purpose or setting.

Psychometric Properties.

In evaluating the evidence base for a specific measurement tool, it is important to note the populations and purposes for which the tool has been developed and evaluated. Consider how reasonable is it to use *the particular form(s)* of the measure, for *the specific measurement function(s)* of interest, in *the specific population* with whom you work. It is worth noting that a single study typically only addresses psychometric properties under one specific scenario, so evidence from multiple studies will often form the evidence base to support using any single measure for the diverse functions and with the diverse populations that you may need it to serve.

Forms. Many measures are available in multiple forms or versions (e.g., parent-report vs. teacher-report). The degree to which each form of the tool has been validated to perform specific assessment functions in specific populations varies. Thus, it is important to consider which specific form(s) are discussed when reviewing information about a tool's validity.

Functions. There are several distinct functions for which school personnel may use mental health assessment tools. To facilitate comparison of psychometric properties, we have grouped these functions into two categories that require different types of evidence to support their validity: (1) surveillance and identification, and (2) progress monitoring. Similar to Beidas et al., (2015), we used criteria published by Hunsley and Mash (2008)[1] as the basis for our classification of the psychometric properties of measures listed in Table 1.

[1] For a recent extension of these criteria, as well as a very useful overview of which psychometric properties are most important for specific assessment functions, see Youngstrom et al. (2017)

Surveillance and Identification. We use surveillance to denote assessment of the degree to which mental health problems are present across a population of students (e.g., as part of a needs assessment to guide investment in new mental health programs/resources). Identification refers to assessment in order to detect or verify the presence and severity of mental health symptoms and/or impairment in individual students. Similar evidence is required to demonstrate reliability and validity for either of these functions, so we combine them in Table 1. Note that all forms of all measures included in Table 1 have at least some supporting evidence for surveillance and identification in at least one youth population.

Progress Monitoring. We use progress monitoring to describe measuring the same student(s) multiple times with the same tool and comparing results to assess change (improvement or deterioration) over time. Here, research must demonstrate that a measure is sensitive to changes in symptoms or functioning when it is administered before, during, and/or after treatment. This psychometric property is also sometimes called "responsiveness" (Mokkink et al., 2010). When studies demonstrating a tool's sensitivity to change have only examined change in scores over fixed intervals of time (e.g., pre-treatment to post-treatment), it may be important to consider the length of the time interval(s) studied. Due to their structure and/or reliability, some tools that have worked well to detect large changes occurring over longer periods of time (e.g., after 6 months of weekly therapy), may not be capable of detecting smaller changes or those measured over a shorter interval (Glasgow & Riley, 2013). Alternatively, some newer tools have been specifically designed for the purpose of obtaining frequent measurements over short intervals to track change as it occurs throughout the course of treatment. The developers typically employ sophisticated research designs and statistical techniques such as growth curve analysis to evaluate the average slope or rate of change in scores over the course of

many repeated measurements (e.g., Athay, Kelley, & Dew-Reeves, 2012; Weisz et al., 2011, 2019). With these studies it may be important to consider how frequently the measures were administered (e.g., weekly vs. monthly). Readers can search for new or additional research using the search terms: "name of measure" + "sensitivity to change" or 'responsiveness' or 'treatment'.

Population. We considered three aspects of the target population: *ages/grades,* *culture/demographics,* and *language*. Each of these can affect the way students' mental health symptoms present and/or how the measurement tool should be interpreted or scored. It is best to use tools validated with samples drawn from populations that match your own target population of students (and their parents/teachers) as closely as possible in all three of these aspects.

Ages/grades. Most measures are first validated for use with students in a relatively narrow range of ages or grade levels with additional validation studies coming later (i.e., demonstrating validity in a broader range of ages/grades). If the measure's website does not indicate that the form of a measure that you are interested in using has been validated for the full range of ages/grades of your students, you may use search terms like: "name of measure" + 'elementary school' or 'adolescent' or 'preschool' to identify additional research.

Culture/demographics. There are many dimensions of culture/demographics that can affect a measurement tool's validity (Kazdin, 2005). Besides age/grade, we have found in the course of our review that some of the most commonly studied dimensions are sex, race/ethnicity, socioeconomic status, rurality, and nationality. Other less frequently studied features may also be important, such as immigrant and refugee status, acculturation, sexual orientation, gender identity, or religious background (Alvidrez, Azocar, & Miranda, 1996; Boehmer, 2002; Stewart & Nápoles-Springer, 2000). Measurement tools are often initially validated in samples either intended to broadly represent the general population or conveniently drawn from populations

easily accessible to researchers (often predominantly white, middle class, and urban or suburban; Guyatt, 1993; Okazaki & Sue, 1995). In either case, many minority identities are often omitted or included in proportions too small for studies to determine whether the tool is equally valid for members of those minority populations as for majority populations (Lau, Chang, & Okazaki, 2010; Okazaki & Sue, 1995). Thus, additional studies are often required to determine whether a measure is reliable and valid with additional populations. Search terms that combine "name of measure" + "population of interest" can be used to identify research describing cross-cultural evaluations. We recommend trying a variety of broad and specific terms related to the population of interest. For example, for a measure validated with Korean-American students, searching "Korean" and "Asian" may yield different results, but both may include relevant studies.

Language. Most of the measurement tools included were first developed by US English speaking authors, but many have now been translated into a wide range of other languages. Ideally, translated versions of measures also undergo a process of *linguistic validation*, to confirm that the translated measure has equivalent meaning to native speakers of the new language. While consensus is lacking about the optimal methods for linguistic validation, the process involves carefully selected procedures to check for and correct inconsistencies between the translated and original measures, such as iterative cycles of translation and back translation and/or direct input from members of the target population (e.g., via focus groups or cognitive interviewing; Epstein, Santo, & Guillemin, 2015). After linguistic validation has been performed, the translated version is typically considered a new form of the measurement tool, with additional research required to evaluate its psychometric properties. Author's websites are a common place to find translated versions of measures and information about whether they have

been linguistically and psychometrically validated. If further information is desired, the following search term: "name of measure" + "name of language" can be used.

Population Reference Data. Population reference data, often called "normative" data or "norms," are any data that help guide interpretation of students' scores by allowing comparison of individual scores with those of a larger population. These data may include recommended clinical cutoff scores, score ranges corresponding to severity (e.g., "normal," "mild," "moderate," "severe")[2], or percentile rankings that describe the percent of the population estimated to score at or below each possible score. Normative data can be very helpful in determining cut points for important decisions (e.g., which students may need intervention). Of course, normative data are most useful when the current student population closely resembles the population from which the reference data was derived. For example, some cross-cultural validation studies have found that population reference data can vary across different populations (e.g., recommended clinical cutoff scores may be higher or lower, scores may be associated with different percentile rankings; Braet et al., 2011; Richardson et al., 2010). Thus experts caution against relying too heavily on population reference data to universally interpret all students' scores in the same way (Myers & Winters, 2002), and whenever measurement tools are used for high-consequence assessment procedures (e.g., a screening that determines which students are eligible for a service based solely on whether they score above/below a particular cutoff) it is important for those procedures to be developed based on reference data from populations that are as similar as possible to the student population. When additional information about normative

[2] While recommended clinical cutoff scores and severity ranges are often derived from population reference data, this is not always the case and thus should not be assumed. In some cases, authors may initially publish suggested cutoff scores or score ranges based on theoretical predictions or pilot data (e.g., Becker, Langberg, Vaughn, & Epstein, 2012). Additionally, teacher- or clinician-report measures sometimes include instructions to assign ratings based on subjective judgement about a child's status relative to a broader population (e.g., Children's Global Assessment Scale; Shaffer, 1983). In either case, this information should not be confused with true population reference data as it cannot be applied in the same way.

data is not available on a measure's website, readers can search using the search terms "name of measure" + 'normative' or 'norms' or 'cutoff'.

Practical Characteristics.

While psychometric properties are highly practical features of measures – they determine exactly where, how, and with whom a measurement tool should be used – there are many other practical characteristics worth considering beyond the research demonstrating a measure's validity for specific purposes. These include how burdensome a measure may be to complete or to score, the type(s) and cost of training materials available to help personnel learn how to use the tool, and other features (e.g., availability of supplemental tools to aid scoring, handouts or manuals to educate staff about the measure, online administration options, and complementary resources such as referral protocols for students who score above clinical cutoff scores).

Quick Tips for Efficiently Investigating Characteristics of Measurement Tools

We hope the following tips, learned while conducting this review, will save readers time performing their own investigations into specific characteristics of measurement tools:

- When exploring a measure's website or article abstracts, try using the browser's 'Find' function to quickly search each page for carefully selected word fragments. For example, search for "sensitiv" to quickly determine whether the webpage contains any information about a measure's sensitivity to change.
- When searching for research articles, place quotation marks around the name of the measure, and around other phrases, to limit results to only articles that contain the specific phrase in its entirety. This is particularly useful for narrowing search results.
- If the link to an online resource is not active, try searching in your favorite internet search engine for "name of the measure" + "PDF" (if you are looking for the measure itself) or

"scoring" (if you are looking for scoring information), etc. You may also try going to the root URL of the broken link. For example, if we had previously hosted a measure at http://youthmentalhealth.missouri.edu/Resources/MEASURE.pdf, but later moved it to a different location on our website, you may be able to find a link to the new location from either the root: http://youthmentalhealth.missouri.edu/Resources, or our homepage: http://youthmentalhealth.missouri.edu.

Review of Specific Measurement Tools

Our review of specific measurement tools focused on standardized measures of symptoms and/or functional impairments associated with anxiety, depression, and/or disruptive behavior problems. We included only those measurement tools that are free[3], validated, and practical for routine implementation in school settings. We began with applicable instruments previously included in Beidas and colleagues (2015) review and identified additional measures via web-based searches and following reference trails. Our search terms were numerous and highly varied, but generally followed the template: ['child' or 'youth' or 'adolescent' or 'school'] + ['depression' or 'anxiety' or 'worry' or 'suicide' or 'suicidality' or 'disruptive behavior' or 'conduct disorder' or 'oppositional defiant disorder' or 'ADHD' or 'inattention' or 'mental health' or 'functioning' or 'impairment'] + ['measure' or 'assessment' or 'screening'].

Given our emphasis on practicality, our search strategy differed in two major ways from the methods of more traditional systematic reviews. (1) In addition to using traditional academic databases (e.g., PsycINFO, PubMed, and Google Scholar), we also used popular, general search engines (e.g., Google, Yahoo, etc.) and subsequently looked up scholarly citations for the measures identified. This proved quite fruitful, as it led us to discover a number of other

[3] One included measure, The Ohio Scales, has a nominal one-time fee for clinicians living outside the state of Ohio

relevant, publicly available lists of assessment tools (American Academy of Pediatrics, 2012; Center for School Mental Health, 2015b, 2015a), and complementary resources that are available online to support practitioners using specific measures. (2) We did not attempt to perform an exhaustive search. Instead, similar to the concept of coding to saturation that is often employed in qualitative research, we 'searched to saturation,' performing searches for each mental health domain until we felt that additional searching was unlikely to return any new information.

The findings from our review are presented in Table 1. In total, thirty-seven measurement tools were identified and grouped into four primary domains of mental health concerns: (1) anxiety, (2) depression and suicidality, (3) disruptive behavior, conduct problems, and attention-deficit/hyperactivity disorder (ADHD); and (4) overall mental health. When permissible, copies of each measure, its administration and scoring instructions, and other supplemental resources have also been compiled at: http://youthmentalhealth.missouri.edu/MeasurementTools

Concluding Comments

In preparation of this guide, we have attempted to remain sharply focused on the practical needs of school social workers. We are optimistic that this resource will help readers more easily access and use evidence-based measurement tools in their day-to-day work with students. However, we would also like to offer several cautions to readers as they use this guide.

First, the measures included in Table 1 are far from an exhaustive list of tools that could be appropriate for students. The tools included in our review are almost entirely problem-focused, but student strengths[4] and competencies[5] are also valuable attributes to assess in school services in order to best understand and promote positive development and well-being (e.g.,

[4] See Climie & Henley (2016) for a recent review of the rationale for strengths-based assessment, models for implementation, and existing measurement tools.
[5] See Duckworth & Yeager (2015) for a review of current issues surrounding the application of noncognitive skill measures in school settings. See Zou (2016) for a review of existing measurement tools.

Climie & Henley, 2016; Garcia, 2014). There are also important domains of mental health problems that we did not review due to space limitations (e.g., trauma, substance use). Within the domains reviewed, we excluded several well-validated measures due to monetary cost, and we may have missed a viable, free measurement tool as well. Furthermore, while we focused on standardized measures, direct observation and real-world data are part of EBA as well (e.g., school attendance rates, direct behavior ratings; Christ, Riley-Tillman, & Chafouleas, 2008).

Second, psychometric properties are relative and dimensional in nature. Evidence from validation studies give us increased confidence in the conclusions that can be made when using a measurement tool in a given context, but no amount of evidence can make a tool universally 'valid' for all purposes and contexts. For this reason, we have focused on denoting whether *any* data is available about these properties (e.g., progress monitoring, U.S. population reference data), and hope interested readers will use the provided links and suggested search strategies to determine whether and how that data may be relevant to their needs.

Finally, we acknowledge that many other pragmatic barriers can impede effective use of measurement tools, irrespective of the characteristics of the tool itself. Common examples include unsupportive school leadership, inflexible school policies, or a school climate that engenders resistance to change. Although the strategies for addressing these barriers are beyond the scope of this review, helpful resources are available to guide school social workers through this process (see those provided by the National Center for School Mental Health at: http://csmh.umaryland.edu/The-SHAPE-System/Playbook-Guides/).

References

Alvidrez, J., Azocar, F., & Miranda, J. (1996). Demystifying the concept of ethnicity for psychotherapy researchers. *Journal of Consulting and Clinical Psychology*, *64*(5), 903–908.

American Academy of Pediatrics. (2012). Mental health screening and assessment tools for primary care. In *Addressing Mental Health Concerns in Primary Care: A Clinician's Toolkit*. Retrieved from http://www.aap.org/en-us/advocacy-and-policy/aap-health-initiatives/Mental-Health/Documents/MH_ScreeningChart.pdf

APA Task Force on Evidence-Based Practice with Children and Adolescents. (2008). *Disseminating evidence-based practice for children and adolescents: A systems approach to enhancing care*. Washington, DC.

Athay, M. M., Kelley, S. D., & Dew-Reeves, S. E. (2012). Brief Multidimensional Students' Life Satisfaction Scale—PTPB Version (BMSLSS-PTPB): Psychometric properties and relationship with mental health symptom severity over time. *Administration and Policy in Mental Health and Mental Health Services Research*, *39*(1–2), 30–40. https://doi.org/10.1007/s10488-011-0385-5

Becker, S. P., Langberg, J. M., Vaughn, A. J., & Epstein, J. N. (2012). Clinical utility of the Vanderbilt ADHD Diagnostic Parent Rating Scale comorbidity screening scales. *Journal of Developmental & Behavioral Pediatrics*, *33*(3), 221–228. https://doi.org/10.1097/DBP.0b013e318245615b

Boehmer, U. (2002). Twenty years of public health research: Inclusion of lesbian, gay, bisexual, and transgender populations. *American Journal of Public Health*, *92*(7), 1125–1130.

Braet, C., Callens, J., Schittekatte, M., Soyez, V., Druart, C., & Roeyers, H. (2011). Assessing emotional and behavioural problems with the Child Behaviour Checklist: Exploring the relevance of adjusting the norms for the Flemish community. *Psychologica Belgica*, *51*(3–4), 213.

Center for School Mental Health. (2015a). *CSMH Comparative Review of Free Measures for School Mental Health*. Retrieved from https://saanys.org/wp-content/uploads/2018/07/CSMH-Comparative-Review-of-Free-Measures-for-School-Mental-Health.pdf

Center for School Mental Health. (2015b). *CSMH Summary of Free Assessment Measures*. Retrieved

from http://soundsupportsk12.com/wp-content/uploads/2014/08/Summary-of-Free-Assessment-Measures-And-Google-Doc-Link-to-Measures-Saved.pdf

Christ, T. J., Riley-Tillman, T. C., & Chafouleas, S. M. (2009). Foundation for the development and use of Direct Behavior Rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention*, *34*(4), 201–213. https://doi.org/10.1177/1534508409340390

Climie, E., & Henley, L. (2016). A renewed focus on strengths-based assessment in schools. *British Journal of Special Education*, *43*(2), 108–121. https://doi.org/10.1111/1467-8578.12131

Connors, E. H., Arora, P., Curtis, L., & Stephan, S. H. (2015). Evidence-based assessment in school mental health. *Cognitive and Behavioral Practice*, *22*(1), 60–73. https://doi.org/10.1016/j.cbpra.2014.03.008

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251. https://doi.org/10.3102/0013189X15584327

Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. Journal of Clinical Epidemiology, 68(4), 435–441.

Furr, R. M. (2017). *Psychometrics: an introduction*. Sage Publications.

Garcia, E. (2014). The Need To Address Noncognitive Skills in the Education Policy. *Economic Policy Institute Briefing Paper #386*, 1–36. https://doi.org/10.1007/978-94-6300-591-3_3

Glasgow, R. E., & Riley, W. T. (2013). Pragmatic measures: what they are and why we need them. *American Journal of Preventive Medicine*, *45*(2), 237–243.

Guyatt, G. H. (1993). The philosophy of health-related quality of life translation. *Quality of Life Research*, *2*(6), 461–465. https://doi.org/10.1007/BF00422220

Hunsley, J., & Mash, E. J. (2008). Developing Criteria for Evidence-Based Assessment: An Introduction to Assessments That Work. In *A guide to assessments that work*.

Kazdin, A. E. (2005). Evidence-Based Assessment for Children and Adolescents: Issues in Measurement Development and Clinical Application. *Journal of Clinical Child & Adolescent Psychology*, *34*(3),

548–558. https://doi.org/10.1207/s15374424jccp3403_10

Lau, A. S., Chang, D. F., & Okazaki, S. (2010). Methodological challenges in treatment outcome research with ethnic minorities. *Cultural Diversity and Ethnic Minority Psychology*, *16*(4), 573–580. https://doi.org/10.1037/a0021371

Merikangas, K. R., He, J.-P., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. *Pediatrics*, *125*(1), 75–81. https://doi.org/10.1542/peds.2008-2598

Merikangas, K. R., He, J.-P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., … Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*(10), 980–989.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Mossman, S. A., Luft, M. J., Schroeder, H. K., Varney, S. T., Fleck, D. E., Barzman, D. H., … Strawn, J. R. (2017). The Generalized Anxiety Disorder 7-item scale in adolescents with generalized anxiety disorder: Signal detection and validation. *Annals of Clinical Psychiatry: Official Journal of the American Academy of Clinical Psychiatrists*, *29*(4), 227-234A. https://doi.org/10.1002/prot.23220

Myers, K., & Winters, N. C. (2002). Ten-year review of rating scales. I: Overview of scale functioning, psychometric properties, and selection. *Journal of the American Academy of Child and Adolescent Psychiatry*, *41*(2), 114–122. https://doi.org/10.1097/00004583-200202000-00004

Okazaki, S., & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. *Psychological Assessment*, *7*(3), 367–375. https://doi.org/10.1037/1040-3590.7.3.367

Richardson, L. P., McCauley, E., Grossman, D. C., McCarty, C. A., Richards, J., Russo, J. E., … Katon, W. (2010). Evaluation of the Patient Health Questionnaire-9 item for detecting major depression

among adolescents. *Pediatrics*, *126*(6), 1117–1123. https://doi.org/10.1542/peds.2010-0852

Shaffer, D. (1983). A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry*, *40*(11), 1228–1231. https://doi.org/10.1001/archpsyc.1983.01790100074010

Stewart, A. L., & Nápoles-Springer, A. (2000). Health-related quality-of-life assessments in diverse population groups in the United States. *Medical Care*, *38*(9), II-102-II–124.

Tam, H. E., & Ronan, K. (2017). The application of a feedback-informed approach in psychological service with youth: Systematic review and meta-analysis. *Clinical Psychology Review*, *55*, 41–55.

Thompson, A. M. (2014). A randomized trial of the Self-Management Training and Regulation Strategy for disruptive students. *Research on Social Work Practice*, *24*(4), 414–427. https://doi.org/10.1177/1049731513509691

Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S. K., … Hoagwood, K. E. (2011). Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *Journal of Consulting and Clinical Psychology*, *79*(3), 369–380. https://doi.org/10.1037/a0023307

Weisz, J. R., Vaughn-Coaxum, R. A., Evans, S. C., Thomassin, K., Hersh, J., Lee, E. H., … Mair, P. (2019). Efficient monitoring of treatment response during youth sychotherapy: The Behavior and Feelings Survey. *Journal of Clinical Child and Adolescent Psychology*, 1–15. https://doi.org/10.1080/15374416.2018.1547973

Youngstrom, E. A., Van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M. L., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, *24*(4), 331–363. https://doi.org/10.1111/cpsp.12207

Zou, K. (2016). *Non-cognitive skills: definitions, measurement and malleability*. *Paper commissioned for the Global Education Monitoring Report 2016, Education for people and planet: Creating sustainable futures for all*. Retrieved from http://unesdoc.unesco.org/images/0024/002455/245576e.pdf

Table 1: Measurement Tools

Measurement Tool		Psychometric Properties[a]				Practical Characteristics			
Mental Health Domain Name of Tool	Where to Obtain Tool and Training	Form and Ages or Grades[b]	Functions[c]	Diverse Pop.[d]	U.S. Ref.[e]	# of Items	Time (minutes)[f]	Training[g]	Additional Features[h]
Anxiety									
Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS)	https://iocdf.org/professionals/training-institute/btti/pediatric-materials/	CIP, CIY (ages 6–17); P, Y (ages 8–17)	SI, PM (3 weeks)	S	S	10	NF	MF	D, FL
Generalized Anxiety Disorder-7 (GAD-7)	https://www.phqscreeners.com/	Y (ages 12–18+)	SI	NF	S	7	AST: <5	MF*	D, FL
Penn State Worry Questionnaire for Children (PSWQ-C)	https://www.childfirst.ucla.edu/resources/	Y (ages 7–17)	SI	S	S	14	AT: <5	MF	FL
Revised Children's Anxiety and Depression Scale (RCADS)	https://www.childfirst.ucla.edu/resources/	P, Y (grades 3–12)	SI, PM (8 weeks)	S	S	25–47	AT: <5–10	MF	D, FL, SC
Screen for Child Anxiety Related Emotion Disorders (SCARED)	https://www.pediatricbipolar.pitt.edu/resources/instruments	P, Y (ages 8–18)	SI, PM (12 weeks)	S	S	41	AT: <10	MF	D, FL, SC

[a] Citations for articles reviewed to evaluate each tool's psychometric properties are available from the corresponding author upon request

[b] CIP = Clinician Interview - Parent/caregiver, CIY = Clinician Interview - Youth/student, CO = Clinician Observation, CR = Clinician-report, P = Parent/caregiver-report, Y = Youth/Student self-report, T = Teacher-report; multiple age/grade ranges separated by commas indicate that separate forms exist for each range

[c] Some evidence available to support use for the following functions: SI = Surveillance and Identification, PM (#) = Progress Monitoring (shortest time interval or most frequent administration schedule for which there is evidence of sensitivity to change for at least one form of the measure)

[d] Diverse populations; S = Some evidence available to support validity of the original and/or an adapted form of the measure in at least one other culture/demographic population beyond white youths in U.S, NF = No evidence found

[e] U.S. population reference data; S = Some population reference data derived from a U.S. sample is available; NF = No U.S. population reference data found

[f] AT = Administration time, ST = Scoring time, AST = Combined time for administration and scoring, NF = No U.S. population reference data found

[g] MF = Manual/written instructions – Free, IF = Interactive online course – Free, VF = Video/recorded webinar – Free, LF = Live (in-person or online) – Free, MP = Manual/written instructions – Proprietary, NF = No training materials found

[h] CR = Complementary Resources (optional add-on components, materials for educating staff, students, or community, etc.), D = Diagnostically relevant (items map on to DSM diagnostic criteria for a specific mental disorder), FL = Foreign Language translation(s), OAP = Online Administration option – Proprietary, OAF = Online Administration option – Free, SC = Scoring tool(s) - Computer program/file, SH = Scoring tool(s) – Hand-scoring/Paper, SO = Scoring tool(s) – Online

Measurement Tool		Psychometric Properties[a]				Practical Characteristics			
Mental Health Domain Name of Tool	Where to Obtain Tool and Training	Form and Ages or Grades[b]	Functions[c]	Diverse Pop.[d]	U.S. Ref.[e]	# of Items	Time (minutes)[f]	Training[g]	Additional Features[h]
Spence Children's Anxiety Scale (SCAS)	http://www.scaswebsite.com/	P (ages 3–6, 6–18); Y (ages 8–18)	SI, PM (13 sessions)	S	S	34–44	AT: <10	MF	D, FL, SC
Depression and s*uicidality*									
Center for Epidemiologic Studies Depression Scale for Children (CES-DC)	https://www.brightfutures.org/mentalhealth/pdf/tools.html	Y (grades 4–12)	SI	S	S	20	AST: 5–10	MF	FL
Columbia Depression Scale (CDS; formerly DISC Depression Scale)	http://www.thereachinstitute.org/images/pdfs/glad-pc-toolkit-2018.pdf	P, Y (ages 11–18+)	SI	S	S	22	AT: 8	MF	CR, D, FL
Columbia-Suicide Severity Rating Scale (C-SSRS)	http://cssrs.columbia.edu/	CIY (ages 5–18+)	SI, PM (6 weeks)	S	S	6–20	NF	IF, LF, MF, VF	CR, FL
Depression Self Rating Scale for Children (DSRSC)	http://www.childrenandwar.org/measures/	Y (ages 8–18)	SI	S	S	18	AST: 5–10	MF	FL
Kutcher Adolescent Depression Scale (KADS)	http://www.shared-care.ca/toolkits-mood	Y (ages 12–17)	SI, PM (weekly)	S	S	6–11	AST: <6	MF	FL
Patient Health Questionnaire Depression Screeners (PHQ-9, PHQ-2)	https://www.phqscreeners.com/; http://www.cqaimh.org/pdf/tool_phq2.pdf	Y (ages 13–18+)	SI	S	S	2–9	AST: <2–5	MF**	D, FL
PHQ-9 Modified for Adolescents/Teens	http://www.gladpc.org/	Y (ages 11–17)	SI	S	S	9	AST: <6	MF	D, FL
Mood and Feelings Questionnaire (MFQ)	https://devepi.duhs.duke.edu/measures/the-mood-and-feelings-questionnaire-mfq/	P, Y (ages 6–18+)	SI, PM (8 weeks)	S	S	13–34	AST: <5	MF	D, FL
Disruptive behavior, conduct problems, and ADHD									
ADHD Rating Scale-IV (ADHD-RS-IV)	http://pcptoolkit.beaconhealthoptions.com/adhd/; https://openlibrary.org/books/OL374088M/ADHD_rating_scale-IV	CR (ages 6–15); P, T (ages 5–18)	SI, PM (5 weeks)	S	S	18	AT: 10	MP	D, FL

Measurement Tool		Psychometric Properties[a]				Practical Characteristics			
Mental Health Domain Name of Tool	Where to Obtain Tool and Training	Form and Ages or Grades[b]	Functions[c]	Diverse Pop.[d]	U.S. Ref.[e]	# of Items	Time (minutes)[f]	Training[g]	Additional Features[h]
Child and Adolescent Disruptive Behavior Inventory (CADBI) Screener	http://measures.earlyadolescence.org/measures/view/40/	P, T (ages 3–16)	SI	S	NF	25	AT: 5–10	NF	D, FL
Disruptive Behavior Disorder Rating Scale (DBDRS)	https://ccf.fiu.edu/about/resources/	P, T (ages 5–10)	SI	S	NF	45	AST: 5–10	MF	D
Impairment Rating Scales (IRS)	https://ccf.fiu.edu/about/resources	P, T (ages 4–12)	SI	NF	S	7	AT: <10	NF	D, FL
NICHQ Vanderbilt Assessment Scales	https://www.nichq.org/resource/caring-children-adhd-resource-toolkit-clinicians	P, T (ages 5–15)	SI, PM (6 months)	S	S	26–55	AST: 10	MF	CR, FL
Modified Overt Aggression Scale (MOAS)	http://www.thereachinstitute.org/images/MOAS.pdf	CIP (ages 6–18+)	SI, PM (5 weeks)	NF	NF	16	AST: 10–15	NF	NF
Overt Aggression Scale (OAS)	https://www.researchgate.net/publication/19220225_The_Overt_Aggression_Scale_for_the_Objective_Rating_of_Verbal_and_Physical_Aggression	CO (ages 5–11)	SI, PM (5 minutes)	NF	NF	27	AT: 5	MF	NF
Swanson, Nolan, and Pelham rating scale (SNAP-IV)	http://www.shared-care.ca/toolkits-adhd	P, T (ages 6–18)	SI, PM (5 weeks)	S	S	18–90	AST: 5–10	MF	D, FL, SH
Strengths and Weaknesses of ADHD symptoms and Normal behavior (SWAN)	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4618695/; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4671522/	P, T (ages 3–18)	SI	S	NF	18–30	AST: 10	NF	D, FL
Overall mental health									
Baby Pediatric Symptom Checklist (BPSC)	https://www.floatinghospital.org/The-Survey-of-Wellbeing-of-Young-Children/Parts-of-the-SWYC/BPSC	P (ages 0–1.5)	SI	S	S	12	AT: <10	MF, VF	CR, FL, SC
Brief Impairment Scale (BIS)	https://www.jaacap.org/article/S0890-8567(09)61660-9/abstract	CIP (ages 4–17)	SI, PM (yearly)	S	S	23	AT: 3–5	MF	FL

Measurement Tool		Psychometric Properties[a]				Practical Characteristics			
Mental Health Domain Name of Tool	Where to Obtain Tool and Training	Form and Ages or Grades[b]	Functions[c]	Diverse Pop.[d]	U.S. Ref.[e]	# of Items	Time (minutes)[f]	Training[g]	Additional Features[h]
Behavior and Feelings Survey (BFS)	https://weiszlab.fas.harvard.edu/measures	P, Y (ages 7–15)	SI, PM (weekly)	NF	NF	12	AT: <1	NF	NF
Children's Global Assessment Scale (CGAS)	http://www.excellenceforchildandyouth.ca/resource-hub/measure-profile?id=77	CR (ages 4–16)	SI, PM (2 weeks)	S	NF	1	AT: 20–30	MF	FL, SC
Columbia Impairment Scale (CIS)	http://www.dhs.state.il.us/page.aspx?item=64974	P, Y (ages 9–17)	SI, PM (6 months)	S	S	13	AT: 5	NF	FL
Early Childhood Screening Assessment (ECSA)	https://medicine.tulane.edu/centers-institutes/tecc/provider-resources/general-screens	P (ages 1.5–5)	SI	S	S	22–40	AT: 5–10, ST: <1	MF	FL
The Ohio Scales	https://sites.google.com/site/ohioscales/	CR, P (ages 5–18); Y (ages 12–18)	SI, PM (3 months)	NF	S	48	AT: 15	MF, VF	CR, FL, OAP, SC
Patient Health Questionnaire for Adolescents (PHQ-A)	https://hhs.texas.gov/doing-business-hhs/provider-portals/health-services-providers/texas-health-steps/forms	Y (ages 13–18)	SI	NF	S	67	AT: 5	MF	D
Symptoms and Functioning Severity Scale (SFSS)	https://peabody.vanderbilt.edu/docs/pdf/cepi/ptpb_2nd_ed/PTPB_2010_Entire_Manual_UPDATE_31212.pdf	CR, P, Y (ages 11–18)	SI, PM (every other week)	S	S	13–27	AT: 5–8	MF	CR, FL, SH
Pediatric Symptom Checklist (PSC and PSC-Y)	https://www.massgeneral.org/psychiatry/services/treatmentprograms.aspx?id=2008	Y (ages 9–18); P (ages 3–17)	SI, PM (3 months)	S	S	17–35	AT: <5, ST: 1–2	MF	FL, OAF, SO
Preschool Pediatric Symptom Checklist (PPSC)	https://www.floatinghospital.org/The-Survey-of-Wellbeing-of-Young-Children/Parts-of-the-SWYC/PPSC	P (ages 1.5–5)	SI	S	S	18	AT: <10	MF, VF	CR, FL
Strengths and Difficulties Questionnaires (SDQ)	http://www.sdqinfo.org	P, T (ages 2–4, 4–10, 11–17, 18+); Y (ages 11–17, 18+)	SI, PM (weekly)	S	S	25–34	AST: 10	MF	CR, FL, OAP, SH

Measurement Tool		Psychometric Properties[a]				Practical Characteristics			
Mental Health Domain Name of Tool	Where to Obtain Tool and Training	Form and Ages or Grades[b]	Functions[c]	Diverse Pop.[d]	U.S. Ref.[e]	# of Items	Time (minutes)[f]	Training[g]	Additional Features[h]
Student Risk Screening Scale (SRSS)	http://www.ci3t.org/screening#srss	T (grades pre-K–12)	SI	S	S	12	AST: 15–20 per class	MF, VF	CR, SC
Top Problems Assessment (TPA)	https://weiszlab.fas.harvard.edu/measures	CIP, CIY (ages 7–13)	SI, PM (weekly)	S	NF	3	AT: 5–10	MF	NF

*Manual includes cutoff scores optimized for adult samples, which may not be optimal for youth. In a clinical sample of 40 adolescents, Mossman et al. (2017) found that scores of 11 and 17 on the GAD-7 were optimal cut points for detection of moderate and severe anxiety, respectively.

**The manual includes cutoff scores optimized for adult samples, which may not be optimal for youth. In a sample of 442 adolescents, Richardson et al. (2010) found that a score of 11 on the PHQ-9 was the optimum cut point for detection of DSM-IV major depression.