

# Chapter 8

## Current Issues and Future Directions in Mental Health Screening

Throughout this volume, we have presented current research and practice in mental health screening in schools as it presently stands. A corpus of prior work is beginning to form supporting the benefits of screening to early prevention and intervention work in schools; however, much remains unknown, even concerning many of the issues we have touched upon in this book. More research must be done in order to ensure that sound science guides the increasingly popular practice of screening children for behavioral and emotional problems in order to avoid untoward outcomes and maximize the benefits of the screening effort.

### Assessing Criterion Validity of Screening Instruments

In Chapter 4, we introduced the reader to a multitude of specific and broadband behavior-rating scales and systems that have been used for mental health screening in schools. When using any screener, it is important that the screener is related to a particular outcome of interest; this is known as criterion validity. In reviewing the literature, researchers have utilized a number of outcome measures when assessing the criterion validity of different screeners. Although this research has yet to compare multiple screeners on the same criteria, initial measurement validation is also critical to current and future efforts in screening research. August et al. (1992) focused on functional impairment using measures of behavioral (using the Internalizing and Externalizing dimensions of the CBCL-PRF (Child Behavior Checklist-Parent Report Form; Achenbach 1991), social (using various subscales of the Walker-McConnell Scale of Social Competence and School Adjustment; Walker and McConnell 1995), and academic (using Woodcock-Johnson Tests of Achievement; McGrew and Woodcock 2001) adjustment. In this way, they could assess impairment independent of clinical diagnoses. An individual found to have functional

impairments in any of these three areas would be positively identified when performing the ROC (Receiver Operating Characteristic) curve analysis.

Other possible criterion/outcome measures include other DSM-IV/DSM-V (Diagnostic and Statistical Manual of Mental Disorders-Fourth/Fifth Edition) (APA 1994, 2013) or ICD (International Classification of Diseases) (WHO 1993) diagnoses (Tarnopolsky et al. 1979; Winter et al. 1999), special education classifications, longer, validated parent and teacher behavior rating scales (Simonian and Tarnowski 2001), mental health referral and treatment histories (Saunders and Wojcik 2004), clinician or teacher-rated levels of impairment (Kelleher et al. 1999; Pagano et al. 2000; Saunders and Wojcik 2004), as well as diagnostic structured interviews such as the Structured Clinical Interview for DSM-IV (SCID-IV; Kobak et al. 1997; Leon et al. 1999; Pagano et al. 2000; Schmitz et al. 1999).

As McFall (2005, p. 318) explained “Only when both sides of the assessment equation have been nailed down is it possible to evaluate what, if anything, the total assessment effort has revealed. Unfortunately, criterion assessment has not received the attention to date that it requires.” No “gold standard” presently exists in psychological assessment research (August et al. 1992; McFall 2005) and all commonly used criterion measures described previously have significant limitations.

For example, obtaining teacher’s ratings of students as the criterion often leads one to suspect method variance since teachers are usually the respondent on the screening measure as well. Special education placement, another commonly used outcome measure, is of unknown reliability and validity and has been found to be determined by factors other than a child’s academic performance or behavior in school, including a child’s sex or race/ethnicity. Kim and Rowe (2004) found that children in special education and those in regular education had identical teacher ratings of their behavior, thus raising the question of why one child was “placed” and another was not (Kim and Rowe 2004). Lastly, the use of DSM diagnosis as an outcome measure is a common practice in psychological assessment literature; however, many of the diagnostic categories found in the DSM have yielded inter-diagnostician reliability estimates lower than the internal consistency estimate of most psychological assessment instruments (i.e., .97; Kamphaus and Frick 2002).

Due to the limitations of all outcomes measures, researchers must emphasize the need for replication when conducting this type of research. Researchers often recommend a “bootstrapping” approach, meaning that we must continually validate new measures against known inferior measures until enough evidence is accumulated to demonstrate that the new measure is superior. In this regard, Schmidt et al. (2004) have observed, “...philosophers of science have shown that it is possible to start with fallible indicators and gradually improve on them, simultaneously refining assessment of the construct (Meehl 1992, p. 141).” However, due to the practical limitations of most applications in schools, typically one measure has been used in isolation in each screening effort; therefore, little data exist comparing screening assessments on important criteria such as false positives, false negatives, and important outcomes following screening.

## Assessing Consequential Validity

Ultimately, evidence of validity based on the consequences of screening will be necessary to defend its use. In order to be cost-effective (an issue described at greater length later in this chapter), early identification must lead to better behavioral and emotional outcomes for children than would be expected in the case of current typical service identification practices. Research must be done to assess whether the intended consequences of such a screening program come to fruition. In order to do this, it would be necessary to implement a school-wide or even district-wide screening program and evaluate the actual consequences longitudinally across a number of years. As described in Chapter 7, some districts are beginning to screen universally; our hope is that these districts continue this effort and allow their data to be used for research that could help to better inform future screening programs.

First, it would be important to determine that the screening instrument was being utilized as intended in order to avoid unintended consequences that could be associated with a positive screen. Due to the ease of administration, schools may be tempted to use the screener as a diagnostic tool rather than as an indication of possible risk. Screeners are inherently less broad-based, assessing a necessarily limited range of behaviors and emotions due to their shorter length. Moreover, in choosing cutoff scores for screeners, we allow more false positives since the screener is only supposed to be the first gate of a multiple-gate system of identification. Therefore, placing too much weight on a positive screen could be quite costly, both in terms of use of resources and mislabeling of children based on preliminary results.

Second, it is critical to make sure that certain children are not being identified more or less often than others as being at-risk for behavioral and emotional maladjustment. For example, if the screener is only identifying those children with externalizing problems and not those with internalizing problems such as anxiety or depression, then those children with internalizing problems would fail to receive the more comprehensive assessment and services that they potentially need. Past research has found that self-reports are especially useful in the identification of students who need intervention for less visible difficulties such as internalizing problems (Merrell et al. 2002). Future research should continue to examine the accuracy and rates of identification of children with various patterns of symptomology, particularly when the goal is broad-based screening for risk for a variety of behavioral or emotional difficulties.

It is also important to determine whether screening instruments overidentify children of specific demographics, including race and gender. In some instances, demographic characteristics have been found to predict special education placement better than academic performance or socioeconomic status (e.g., Hosp and Reschly 2004; Skiba et al. 2008). In terms of race, African American students are over-represented in the Emotional Disturbance (ED) category of special education, most

specifically, and in special education, more broadly (Ahram et al. 2011; Hosp and Reschly 2003; Jasper and Bouck 2013; MacMillan and Reschly 1998; Skiba et al. 2008). Males are also overrepresented in special education at a ratio of between 1.5:1 and 3.5:1 (Christenson et al. 1983; Coutinho and Oswald 2005), and are particularly at-risk for being referred to special education for behavioral problems (Bryan et al. 2012; Coutinho and Oswald 2005; Wallace et al. 2008). Therefore, if one of the goals of screening is to provide information prior to a special education referral, it is imperative that the screening instrument does not exacerbate, or even better, begins to ameliorate, current patterns of disproportionality associated with being referred to and ultimately receiving special education services for behavioral and emotional problems.

Lastly, we would want to evaluate how the children who are identified by the screening instrument as being at-risk are being served. What types of early interventions are in place in schools, and are these interventions addressing the needs that are being identified by the screening process and follow-up assessments? As Goodman et al. (2003) explained, “There would obviously be no point in identifying a greater proportion of children with psychiatric disorders in the community if the only consequence were greater access to ineffective treatments...even if treatments are effective, there is no point identifying more children in need of treatment if existing services are already overstretched and no resources are available” (p. 171).

In our own work in schools, linking screening results with appropriate interventions has admittedly been one of the most difficult challenges we have faced. In many instances, it is necessary to work within the intervention framework that already exists within the schools, as many districts lack the time and resources required to develop new interventions based on screening results. However, it is also true that very little research has been conducted regarding the best steps to take following a positive screening, including intervention decisions. As stated by Vannest (2012), following a screening effort the school must make decisions regarding whom to serve, when to serve them, what service to provide, and who will provide that service. Vannest et al. (2008) compiled a compendium of empirically-validated interventions for various behavioral and emotional challenges; however, teacher training is often necessary, many interventions may be useful for only one specific difficulty, and schools may find it difficult to address all of the needs of children simultaneously. More research is needed concerning matching interventions to specific difficulties, focusing on progress over time. In addition, research on the impact of group interventions or school-wide prevention efforts on student outcomes over time could be useful for providing schools with limited resources with more efficient alternatives to address the needs of the students. This type of longitudinal research is costly and time-intensive; however, it is a crucial step in evaluating the long-term effectiveness of a universal screening program.

## Assessing the Usefulness of Multiple-Gated Systems and Available Informants

In Chapter 6, we discussed the available research concerning multiple-gated approaches to screening, as well as choosing an informant for the screening assessment. Presently, there is no consensus regarding either the best number of gates, or the procedures to be implemented at each gate. Although some systems use three gates (e.g., Systematic Screening for Behavior Disorders (SSBD); Walker and Severson 1992), others have suggested two gates are sufficient (e.g., VanDeventer 2008; see Chapter 6 for a full review of the literature on this issue). Longitudinal studies of multiple-gated screening procedures are particularly crucial for determining the false negative and false positive rates of each configuration of gates to identify children who later develop significant behavioral and emotional problems.

Similarly, the best use of various informants, whether in a multiple-gated system or individually, when screening is still undecided. As reviewed in Chapter 6, the choice of informant(s) may depend upon the behavior of interest, as well as the age of the child and other characteristics of the potential informant(s). For example, teachers and parents may be best suited to provide information about younger children, particularly regarding their externalizing behaviors (Loeber et al. 1990). However, older children and adolescents may be the best source of information concerning their own internalizing problems (Loeber et al. 1991; Pagano et al. 2000; Smith 2007; Youngstrom et al. 2000).

Realistically, decisions regarding the number of gates and choice of informants are often driven by feasibility within a given setting. For example, among children who are just beginning at a new school setting (e.g., kindergarten), asking parents to complete screening forms at a registration event could serve as an efficient method of gathering information universally. However, for students in the 11th grade, it might be more difficult to send forms home to parents and have them successfully completed and returned to the school; in that instance, teacher- or student-reports completed on-site often yield a higher response rate and therefore, decrease the likelihood of students who are in need of follow-up assessment from “falling through the cracks” at the screening stage. As described in Chapter 7, student-report screening among middle and high school students may be more acceptable to teachers, as students often have multiple teachers in these grades, which makes identifying the “best” teacher informant a practical challenge.

To be most successful, practice must inform future research as much as research must inform future practice regarding the feasibility of collecting information over multiple gates, and/or from multiple informants. Researchers must seek out school districts who are already using multiple- or single-gated approaches to screening to better understand the implementation challenges that schools face when undertaking a screening effort (e.g., Dever et al. 2012). Even the strongest of research findings is not useful if it has limited external validity or generalizability to a real-world

situation of the actual practice of screening in schools; therefore, the development of future best practices should include strong partnerships between researchers and school districts. Only with such applied work can researchers begin to understand the feasibility and utility of multiple-gated systems and the number and types of informants that are both necessary and plausible to involve in the process.

## **Assessing the Stability of Screening Results over Time**

When embarking upon a universal screening program, it is important to consider a long-term plan for integrating screening into the procedures of a school or district. One decision that is frequently overlooked is how often the school or district will screen. Screening data are excellent for providing a snapshot of risk for behavioral and emotional problems at one point of time. However, these sorts of issues are often fluid and may depend upon individual and contextual circumstances at any given time; as such, future research is necessary for determining the frequency with which screening should be conducted in order to maximize classification accuracy, minimize false negatives, and reduce costs (an issue to be discussed further in the next section).

Although previous research suggests that behavioral and emotional problems are fairly stable across time (Essex et al. 2009; Levitt et al. 2007), there is limited empirical knowledge regarding the stability of screening scores and classifications (at-risk vs. not at-risk) over time. Some scholars in the field have recommended that screening should occur three times per year in order to identify students who may need additional services, those whose functioning may have deteriorated over time, and new students who may be in need of support (Parisi et al. 2014; Walker et al. 2014). Walker (2010) suggested that screening for behavioral and emotional risk take place once early in the academic year, with a follow-up screening only for schools with higher student transiency in the early spring in order to avoid missing any new students who have transferred into the school and might be in need of services.

Dowdy and colleagues examined the stability of screening scores and found that behavioral and emotional risk screening classifications were moderately stable across a 4-year period (Dowdy et al. 2014b). Moderate stability coefficients were also found for both overall risk and for domains of risk (internalizing, externalizing, adaptive skills, and school problems) in a district-wide effort across 2 years (Dever et al. *in press*). These empirical studies call into question the need for multiple screenings within the same academic year. However, there is a continued need to examine screening frequency and the stability of screening scores to determine the optimal screening schedule for schools and districts with the simultaneous goals of maximizing efficiency and minimizing risk to students. As service delivery decisions are often made based on these results, the frequency of screening is an important issue to evaluate in future research.

## Assessing the Cost/Benefit Ratio

When evaluating potential screening tools, the accuracy of identification is critical to gathering good data; however, in the context of actual schools, the practicality of the instrument is also of utmost importance. One must balance the amount of information needed to reliably identify those children who are at-risk for emotional and behavioral problems against the time and monetary resources of those schools and districts that will be collecting the information. The scientific literature indicates that the impracticality of many screening measures has largely contributed to their lack of adoption on a wide scale in both pediatric and school settings (Flanagan et al. 2003; Saunders and Wojcik 2004; Schmitz et al. 1999). Therefore, one must compare the time and cost of adding additional levels of assessment and informants against the benefits that are gained in terms of increased accuracy of identification with each additional cost. For example, when an informant or gate is eliminated in a screening design, the number of children receiving full diagnostic assessments and more intensive interventions at a later time will likely increase due to the decrease in early identification and prevention procedures.

Researchers must consider what school personnel will tolerate in terms of time and financial investment. Teachers, school psychologists, administrators, and other educational stakeholders are extremely busy and their time is valuable. If the screening process takes too long for teachers to complete, universal screening is unlikely to be adopted successfully. Moreover, the costs associated with implementing a universal emotional and behavioral screening program, including personnel and materials costs, should be thoroughly examined. A cost-benefit ratio between the resources needed to conduct the screening program and the amount of information needed to make accurate predictions is a necessary step to future research.

Before embarking upon a universal screening program, school psychologists and other school personnel commonly express concerns about the “cost” of identifying a large number of students who will be left unable to be served by them (Dever et al. 2012). This concern is understandable, given the plethora of duties with which schools are tasked on a daily basis. However, upon presenting the data back to schools and districts, it has been our experience that most students who are identified as at-risk are already receiving services; therefore, the best approach for these students may be to merely monitor their progress rather than a full reevaluation or change in intervention. According to a population-based model, we should expect approximately 20% of students to be rated at elevated levels of risk; although somewhat anecdotal, we have found that only about 20% of this 20%, or about 4% of total students screened, emerge as new cases that have not been previously identified or are not currently receiving services. For some of these students, a more comprehensive assessment might reveal that the screening result was a false positive; for others, this is a true positive result, and early intervention efforts should be considered before the detected risk worsens.

In general, research is needed to determine whether a universal screening program adds to the burdens that teachers, school psychologists, and others already face or

alleviates burdens of financial expense and time by increasing accuracy of referral and identifying children earlier. Presumably, accurate screening will decrease the need for time- and money-consuming procedures such as special education referral and full evaluations as well as more intensive interventions. However, empirical evidence is needed to support, or refute, this assumption concerning the relative costs and benefits of a universal screening program.

## Addressing Perceptions of Screening

As demonstrated by several high profile legal actions and parental complaints regarding emotional and behavioral screening, public perceptions of screening must be addressed. One concern is that asking questions about suicidal intent may entice adolescents to actively consider suicide when they would have not done so otherwise. Gould et al. (2005) recently evaluated the iatrogenic risk of youth suicide screening programs and found no evidence to suggest that this is the case. In fact, their findings suggest that the screening may have been beneficial for students with symptoms of depression or previous suicide attempts. Although scientific research has not supported the hypothesis of iatrogenic risk of youth suicide screening programs, this public fear represents a significant barrier to public acceptance of universal emotional and behavioral screening of children and adolescents. In August 2003, Illinois was the first state to pass legislation in which a plan, drafted by the *Illinois Children's Mental Health Partnership* (ICMHP), recommended that "all children receive periodic social and emotional developmental screens" (Barlas 2004). This plan was met with great opposition by a group of parents who felt that, according to Barbara Shaw, chairman of the (ICMHP), "the schools have no place futzing with their children's mental health." The parents feared that emotional and behavioral screening would lead to the unnecessary labeling and medicating of their children. These occurrences suggest that public opinion of mental health disorders and our ability to detect and treat them may not yet be at the point where universal emotional and behavioral screening would be widely accepted.

Issues regarding the use of active or passive consent for collecting information on behavioral and emotional risk from students could have an influence on public perception and must be considered prior to the implementation of a universal screening program. Concerns regarding parental rights, student assent, and confidentiality must be addressed explicitly prior to any screening effort. Active consent requires signed, explicit permission from a parent or guardian prior to screening his/her child; passive permission provides the parent or guardian with the opportunity to withdraw his/her child from the screening, with a nonresponse indicative of the passive provision of permission. In the US, it is common for school districts to screen for vision, hearing, and academic concerns routinely with only passive or implied consent procedures in place. However, as Gardner (2011) made clear, behavioral and emotional health screening information may be considered more sensitive than those other domains, both by parents and the local jurisdiction; therefore, in some



circumstances, active consent may be necessary prior to beginning any screening. The impact of an active versus passive consent procedure on response rate, and by extension, hit rate of identification has not yet been examined empirically; future practice could benefit from such an examination of the strengths and limitations of each type of consenting process on a screening effort.

In addition to public perceptions of screening, it is important for future research to collect information about the perceptions of educational stakeholders concerning screening procedures. Social validity refers to teachers' and others' beliefs that the procedures being conducted are both feasible and useful. Some research has found that while teachers perceive screening as useful and acceptable overall, they have concerns regarding feasibility of both universal screening and intervention efforts following screening (Greer et al. 2012). Social validity is critical to the ultimate success of a screening program, as it has been related to fidelity of implementation of school-based programming (e.g., Lane et al. 2009). Therefore, it is imperative for researchers to consider ways to improve social validity of screening work in order for such research to be implemented well in practical settings.

Finally, the buy-in of school leadership is necessary prior to starting a screening program, as the results of screening must be integrated into the procedures of a school or district in order to have any sort of meaningful impact on the school level (Parisi et al. 2014). In the best-case scenario, screening data would be used in a comprehensive data-based decision-making model concerning prevention and intervention for behavioral and emotional problems at the school- and/or district-levels. For this to become a reality, school leaders must be dedicated to the universal screening effort as an iterative process rather than an isolated incident. Past work has found that consensus or near consensus (e.g., 80% or higher) is not necessarily needed among school personnel prior to beginning the screening program, but that a well-executed screening effort with clear results and strong leadership behind it can actually increase buy-in for subsequent screening (Parisi et al. 2014). As school psychologists are often the best-prepared in the areas of assessment, interpretation, and intervention, the commitment of the school psychologist to the screening program is essential (Dever et al. 2012). Despite the knowledge of the importance of school administrators and school psychologists for the success of a universal screening program, there is limited research on *how* to increase the commitment of these stakeholders. In the future, it is imperative that these issues be studied empirically in order to identify areas that could be strengthened or emphasized when introducing a new district or new personnel to the potential of universal screening to improve their students' behavioral, emotional, and ultimately, academic, outcomes.

## Considering the Diversity of the Student Population

The growing diversity of students enrolled in US schools and the globalization of education in general, requires a consideration of how to engage in screening, assessment, and intervention efforts with diverse children and families in a cultur-

ally competent manner. The behavioral and emotional constructs of interest when screening are often influenced by and defined differently within each culture (Yates et al. 2008). Therefore, cultural information must be integrated into the screening process in order to avoid labeling behaviors that are normative in one culture as maladaptive (Dowdy et al. 2014a). In addition, the language abilities of the student being assessed, as well as any family members as informants, must be considered when determining the best way to collect screening information. Dowdy et al. (2014a) have recommended that informants who are English language learners are screened in their native language and preferred modality (i.e., oral vs. written). In addition, the language and reading level of consent forms must be appropriate for the selected informants. When screening among linguistically diverse populations, it is essential that measures are not simply translated from the languages in which they were originally developed to the new language as this may have an unexpected effect on the psychometric properties, meaning, and interpretation of the measure. As such, when designing a screening program the school or district should choose instruments for which there is evidence of appropriateness and psychometric properties within the entire population that is going to be screened (Dowdy et al. 2014a).

When the decision is made to use screening measures with a different population from that on which the measures were originally normed, there must be evidence of measurement invariance for the newly-intended population. In other words, one must ask whether the assessment is still measuring the domains of interest, with the same precision and accuracy, when the assessment is applied to a new group of students. Measurement invariance is important both when translating instruments to a new language, and when using an instrument in the same language, but among a new group of students. For example, if a screening instrument was developed and normed among a sample of students who were in middle school and 95% Caucasian American, a researcher must consider whether that instrument performs the same psychometrically in a high school whose student body is 80% African American, prior to interpreting the results within this new context.

Measurement invariance is emerging as an important venue for continued research in screening. Some researchers have provided preliminary evidence of measurement invariance of omnibus rating scales (e.g., *Child Behavior Checklist (CBCL)*; Gross et al. 2006) and screening instruments (e.g., *BASC-2 Behavioral and Emotional Screening System*; Dowdy et al. 2011; Raines 2011) across language translations, racial and ethnic groups, and gender. However, much more work is needed in this area. Since inferences are made based on the outcome of screening, it is imperative that the instrument of choice is not biased against, or toward, the identification of certain students based on linguistic and cultural characteristics. Otherwise, screening results may serve to perpetuate the disproportionate representation of students that is currently seen in our referral and service systems in schools.

In some instances, practitioners may develop locally-based norms to make screening decisions as opposed to national norms that may not adequately represent the diversity of their own student body (Dowdy et al. 2014a). Local norms have the benefit of identifying the students with the most need in comparison to one's own local population (Glover and Albers 2007). However, national norms have the

benefit of comparing a student's results to similar peers in one's own grade level, age group, or gender group. Research is needed to compare and contrast the results of screening efforts utilizing local and national norms in order to understand the circumstances under which each method might be the most appropriate and useful.

## Conclusions and Final Thoughts

In this volume, we have attempted to compile the existing evidence regarding best practice and recommendations for screening based on empirical research. Although the rich history of mental health screening is clear, the concept of screening universally for risk for behavioral and emotional disorders has just come to the forefront of school psychology research and practice in the past decade or so. In this nascent field, there is much opportunity for further research regarding screening instruments, multiple-gated systems, choice of informants, and other relevant issues related to creating a comprehensive screening-to-intervention system.

Although such a turn-key system is yet to be developed, the field of mental health screening in schools has grown exponentially in the past few years, and this growth is likely to continue. The goals of prevention and early intervention fit well within an RtI framework and make sense intuitively, given the desire to address and ultimately ameliorate behavioral and emotional problems among students as early as possible. Great strides have been made concerning the development of screening instruments, their application in real-world school settings, and their links to intervention efforts; despite this progress, the field of mental health screening in schools remains an area that is "ripe for the picking" among researchers in the area of school psychology. Our hope is that this book will inspire both practitioners and researchers to continue to work toward establishing universal screening for behavioral and emotional risk in schools as typical, rather than exceptional, practice.

## References

- Achenbach, T. M. (1991). *Integrative Guide to the 1991 CBCL/4-18, YSR, and TRF Profiles*. Burlington: University of Vermont, Department of Psychology.
- Ahram, R., Fergus, E., & Noguera, P. (2011). Addressing racial/ethnic disproportionality in special education: Case studies of suburban school districts. *Teachers College Record, 113*(10), 2233–2266.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- August, G. J., Ostrander, R., & Bloomquist, M. J. (1992). Attention deficit hyperactivity disorder: Epidemiological screening method. *American Journal of Orthopsychiatrics, 62*, 387–396.
- Barlas, S. (2004). Illinois passes controversial child screening plan. *Psychiatric Times, 21*. <http://www.psychiatrictimes.com/articles/illinois-passes-controversial-child-screening-plan>. Accessed 26 June 2015.

- Bryan, J., Day-Vines, N. L., Griffin, D., & Moore-Thomas, C. (2012). The disproportionality dilemma: Patterns of teacher referrals to school counselors for disruptive behavior. *Journal of Counseling and Development, 90*(2), 177–190.
- Christenson, S., Ysseldyke, J. E., Wang, J. J., & Algozzine, B. (1983). Teachers' attributions for problems that result in referral for psychoeducational evaluation. *The Journal of Educational Research, 76*, 174–180.
- Coutinho, M. J., & Oswald, D. P. (2005). State variation in gender disproportionality in special education findings and recommendations. *Remedial and Special Education, 26*(1), 7–15.
- Dever, B. V., Raines, T. C., & Barclay, C. M. (2012). Chasing the unicorn: Practical implementation of universal screening for behavioral and emotional risk. *School Psychology Forum, 6*(4), 108–118.
- Dever, B. V., Dowdy, E., Raines, T. C., & Carnazzo, K. (In Press). Stability and change of behavioral and emotional screening scores. *Psychology in the Schools*.
- Dowdy, E., Dever, B. V., DiStefano, C., & Chin, J. (2011). Screening for emotional and behavioral risk among students with limited English proficiency. *School Psychology Quarterly, 26*, 14–26.
- Dowdy, E., Kamphaus, R. W., Twyford, J., & Dever, B. V. (2014a). Culturally competent screening approaches and treatments. In M. Weist, N. Lever, C. Bradshaw, & J. Owens (Eds). *Handbook of School Mental Health* (2nd ed.). New York: Springer.
- Dowdy, E., Nylund-Gibson, K., Felix, E., Morovati, D., Carnazzo, K., & Dever, B. V. (2014b). Long-term stability of screening for behavioral and emotional risk. *Educational and Psychological Measurement, 74*(3), 453–472. doi:10.1177/0013164413513460.
- Essex, M. J., Kraemer, H. C., Slattery, M. J., Burk, L. R., Boyce, W. T., Woodward, H. R., & Kupfer, D. J. (2009). Screening for childhood mental health problems: Outcomes and early identification. *Journal of Child Psychology and Psychiatry, 50*(5), 562–570.
- Flanagan, K. S., Bierman, K. L., & Kam, C. M. (2003). Identifying at-risk children at school entry: the usefulness of multibehavioral problem profiles. *Journal of Clinical Child and Adolescent Psychology, 32*, 396–407.
- Gardner, N. (2011). Potentials and impediments to universal, school-based screening for behavioral and emotional risk: A critical discourse analysis of current case law. (Unpublished doctoral dissertation). Georgia State University, Atlanta, GA.
- Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45*(2), 117–135.
- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2003). Using the strengths and difficulties questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *International Review of Psychiatry, 15*, 166–172.
- Gould, M. S., Marrocco, F. A., Kleinman, M., Thomas, J. G., Mostkoff, K., Cote, J., & Davies, M. (2005). Evaluating iatrogenic risk of youth suicide screening programs. *Journal of the American Medical Association, 293*, 1635–1643.
- Greer, F. W., Wilson, B. S., DiStefano, C., & Liu, J. (2012). Considering social validity in the context of emotional and behavioral screening. *School Psychology Forum, 6*(4), 148–159.
- Gross, D., Fogg, L., Young, M., Ridge, A., Cowell, J. M., Richardson, R., & Sivan, A. (2006). The equivalence of the Child Behavior Checklist/1 1/2–5 across parent race/ethnicity, income level, and language. *Psychological Assessment, 18*(3), 313–323.
- Hosp, J. L., & Reschly, D. J. (2003). Referral rates for intervention or assessment: a meta-analysis of racial differences. *The Journal of Special Education, 37*(2), 67–80.
- Hosp, J. L., & Reschly, D. J. (2004). Disproportionate representation of minority students in special education: Academic, demographic, and economic predictors. *Exceptional Children, 70*(2), 185–199.
- Jasper, A. D., & Bouck, E. C. (2013). Disproportionality among African American students at the secondary level: Examining the MID disability category. *Education and Training in Autism and Developmental Disabilities, 48*(1), 31–40.
- Kamphaus, R. W., & Frick, P. J. (2002). *Clinical assessment of child and adolescent personality and behavior* (2nd ed.). Boston: Allyn & Bacon.

- Kelleher, K. J., Moore, C. D., Childs, G. E., Angelilli, M. L., & Comer, D. M. (1999). Patient race and ethnicity in primary care management of child behavior problems. *Medical Care, 37*, 1092–1104.
- Kim, S., & Rowe, E. (2004, April). Similar risks, dissimilar outcomes: Rethinking risk in educational context. Paper presented at the annual conference of the American Educational Research Association, San Diego, CA.
- Kobak, K. A., Taylor, L. vH., Dottl, S. L., Greist, J. H., Jefferson, J. W., Burroughs, D., Katzelnick, D. J., & Mandell, M. (1997). Computerized screening for psychiatric disorders in an outpatient community mental health clinic. *Psychiatric Services, 48*, 1048–1057.
- Lane, K. L., Kalberg, J. R., Bruhn, A. L., Driscoll, S. A., Wehby, J. H., & Elliott, S. N. (2009). Assessing social validity of school-wide positive behavior support plans: Evidence for the reliability and structure of the primary intervention rating scale. *School Psychology Reviewer, 38*, 135–144.
- Leon, A. C., Kathol, R., Portera, L., Farber, L., Olfson, M., Lowell, K. N., & Sheehan, D. V. (1999). Diagnostic errors of primary care screens for depression and panic disorder. *International Journal of Psychiatry in Medicine, 29*, 1–11.
- Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology, 45*, 163–191.
- Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perceptions of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology, 19*, 136–143.
- Loeber, R., Green, S. M., Lahey, B. B., Stouthamer-Loeber, M. (1991). Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology, 19*, 75–95.
- MacMillan, D. L., & Reschly, D. J. (1998). Overrepresentation of minority students the case for greater specificity or reconsideration of the variables examined. *The Journal of Special Education, 32*(1), 15–24.
- McFall, R. M. (2005). Theory and utility—key themes in evidence-based assessment: Comment on the special section. *Psychological Assessment, 17*, 312–323.
- McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual. Woodcock-Johnson III*. Itasca: Riverside Publishing.
- Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality, 60*(1), 117–174.
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194–216.
- Merrell, K. W., McClun, L. A., Kempf, K. G., & Lund, J. (2002). Using self-report assessment to identify children with internalizing problems: Validity of the internalizing symptoms scale for children. *Journal of Psychoeducational Assessment, 20*(3), 223–239.
- Pagano, M. E., Cassidy, L. J., Little, M., Murphy, J. M., & Jellinek, M. S. (2000). Identifying psychosocial dysfunction in school aged children: The Pediatric Symptom Checklist as a self-report measure. *Psychology in the Schools, 37*, 91–106.
- Parisi, D. M., Ihlo, T., & Glover, T. A. (2014). Screening within a multitiered early prevention model: Using assessment to inform instruction and promote students' response to intervention. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 19–46). Washington, DC: American Psychological Association.
- Raines, T. C. (2011). Screening as the great equalizer: Eliminating disproportionality in special education referrals. (Unpublished doctoral dissertation). Georgia State University, Atlanta, GA.
- Saunders, S. M. & Wojcik, J. V. (2004). The reliability and validity of a brief self-report questionnaire to screen for mental health problems: the health dynamics inventory. *Journal of Clinical Psychology in Medical Settings, 11*, 233–241.
- Schmidt, N. B., Kotov, R., & Joiner, T. E. (2004). *Taxometrics: Toward a new diagnostic scheme for psychopathology*. Washington, DC: American Psychological Association.

- Schmitz, N., Kruse, J., Heckrath, C., Alberti, L., & Tress, W. (1999). Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Social Psychiatry and Psychiatric Epidemiology*, *34*, 360–367.
- Simonian, S. J., & Tarnowski, K. J. (2001). Utility of the pediatric symptom checklist for behavioral screening of disadvantaged children. *Child Psychiatry and Human Development*, *31*, 269–278.
- Skiba, R. J., Simmons, A. B., Ritter, S., Gibb, A. C., Rausch, M. K., Cuadrado, J., & Chung, C. (2008). Achieving equity in special education: History, status, and current challenges. *Exceptional Children*, *74*(3), 264–288.
- Smith, S. R. (2007). Making sense of multiple informants in child and adolescent psychopathology: A guide for clinicians. *Journal of Psychoeducational Assessment*, *25*, 139–149.
- Tarnopolsky, A., Hand, D. J., McLean, E. K., Roberts, H., & Wiggins, R. D. (1979). Validity and uses of a screening questionnaire (GHQ) in the community. *British Journal of Psychiatry*, *134*, 508–515.
- VanDeventer, M. C. (2008). Universal emotional and behavioral screening for children and adolescents. Unpublished doctoral dissertation.
- Vannest, K. J. (2012). Implementing interventions and progress monitoring subsequent to universal screening. *School Psychology Forum*, *6*(4), 119–136.
- Vannest, K. J., Reynolds, C. R., & Kamphaus, R. W. (2008). *BASC-2: Intervention guide*. Circle Pines: Pearson.
- Walker, B. A. (2010). Effective schoolwide screening to identify students at risk for social and behavioral Problems. *Intervention in School and Clinic*, *46*, 104–110.
- Walker, H. M., & McConnell, S. R. (1995). *Walker-McConnell scale of social competence and school adjustment, elementary version: User's manual*. San Diego: Singular Publishing Group, Inc.
- Walker, H. M., & Severson, H. H. (1992). *Systematic screening for behavior disorders (SSBD)*. Longmont: Sopris West.
- Walker, H. M., Small, J. W., Severson, H. H., Seeley, J. R., & Feil, E. G. (2014). Multiple-gating approaches in universal screening within school and community settings. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 47–75). Washington, DC: American Psychological Association.
- Wallace, J. M., Goodkind, S., Wallace, C. M., & Bachman, J. G. (2008). Racial, ethnic, and gender differences in school discipline among U.S. high school students: 1991–2005. *Negro Education Review*, *59*(1–2), 47–62.
- Winter, L. B., Steer, R. A., Jones-Hicks, L., & Beck, A. T. (1999). Screening for major depression disorders in adolescent medical outpatients with the Beck Depression Inventory for primary care. *Journal of Adolescent Health*, *24*, 389–394.
- World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. World Health Organization. Geneva: Author.
- Yates, T., Ostroskly, M. M., Cheatham, G. A., Fettig, A., Shaffer, L., & Santos, R. M. (2008). Research synthesis on screening and assessing social-emotional competence. The Center on the Social and Emotional Foundations for Early Learning. Retrieved from [Vanderbilt.edu/csefel](http://Vanderbilt.edu/csefel).
- Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology*, *68*, 1038–1050.