

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Protective Factor Screening for Prevention Practice: Sensitivity and Specificity of the DESSA-Mini

### Permalink

<https://escholarship.org/uc/item/32j6r9bq>

### Journal

School Psychology, 32(4)

### ISSN

2578-4218

### Authors

Shapiro, Valerie B  
Kim, BK Elizabeth  
Robitaille, Jennifer L  
[et al.](#)

### Publication Date

2017-12-01

### DOI

10.1037/spq0000181

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

**Protective Factor Screening for Prevention Practice: Sensitivity and Specificity of the DESSA-Mini**

Valerie B. Shapiro, B. K. Elizabeth Kim, Jennifer L. Robitaille, and Paul A. LeBuffe Online First Publication, October 13, 2016.  
<http://dx.doi.org/10.1037/spq0000181>

CITATION

---

Shapiro, V. B., Kim, B. K. E., Robitaille, J. L., & LeBuffe, P. A. (2016, October 13). Protective Factor Screening for Prevention Practice: Sensitivity and Specificity of the DESSA-Mini. *School Psychology Quarterly*. Advance online publication.  
<http://dx.doi.org/10.1037/spq0000181>

## Protective Factor Screening for Prevention Practice: Sensitivity and Specificity of the DESSA-Mini

### Abstract

The Devereux Student Strengths Assessment Mini (DESSA-Mini; Naglieri, LeBuffe, & Shapiro, 2011/2014) was designed to overcome practical obstacles to universal prevention screening. This article seeks to determine whether an entirely strength-based, 8-item screening instrument achieves technical accuracy in routine practice. Data come from a district-wide implementation of a new social emotional learning (SEL) initiative designed to promote students' social-emotional competence. All students, kindergarten through Grade 8, were screened using the DESSA-Mini. A random 5 students per classroom received additional assessment. Concurrent and predictive criterion studies were conducted using the full DESSA as well as administrative records of serious disciplinary infraction. The DESSA-Mini showed excellent internal reliability, exceeding .90. Negligible to small differences were found between scores on the DESSA-Mini screen and the DESSA full assessment. Classification consistency between the DESSA-Mini and the DESSA was high (87%–94%) in routine practice, with sensitivity and specificity estimates exceeding Glascoe's (2005) standards. Finally, predictive validity of the DESSA-Mini was reliable; students screened as having a Need for SEL Instruction at the beginning of the year were 4.5 times more likely to have a record of serious disciplinary infraction at the end of the school year compared with those who were not identified ( $p < .001$ ). These findings compare quite favorably with other instruments used in schools to screen entire student populations, in cases where such analyses have been conducted, and is consistent with a practice preference of identifying, but not over-identifying, students for accelerated preventative interventions for mental, emotional, and behavioral problems.

Author Notes: This project, and the associated data collection and analyses, was enabled through the support of the Pennsylvania Departments of Education, Health, and Public Welfare, and the Hellman Foundation. The opinions expressed in this article do not reflect the views of any funder or host organization. A version of this article was presented at the 2016 American Education Research Association Annual Conference in Washington, D.C. The authors wish to acknowledge all of the teachers, staff, and administrators who participated in this project, and specifically Gerald L. Zahorchak, C. Russell Mayo, Susan Lozada, Robin Powlus, and John Monahan. The authors also wish to thank Katherine Ross, Julie Koenigsberg, and Jennifer Croner, who each played an essential role in the completion of this work. Although some authors were involved in the development of the DESSA, no author of this article receives any financial remuneration from the sale of the assessments or resources mentioned within this article.

Correspondence concerning this article should be addressed to Valerie B. Shapiro, School of Social Welfare, University of California, Berkeley, 120 Haviland Hall, Berkeley, CA 94720-7400.  
E-mail: [VShapiro@Berkeley.edu](mailto:VShapiro@Berkeley.edu)

Students need social-emotional competence in order to achieve important developmental outcomes, including school success (Domitrovich, Cortes, & Greenberg, 2007; Greenberg et al., 2003; Payton et al., 2008). However, 14% to 20% of school-age children and youth in the United States have a mental, emotional, or behavioral (MEB) disorder that interferes with learning (Doll, 1996; O'Connell, Boat, & Warner, 2009). Research has demonstrated that high-quality, well implemented, social-emotional learning (SEL) programs build social-emotional competence, reduce MEB problems, and promote academic achievement (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011).

Effective SEL programs, however, are not as widely used in schools as one might expect (Gottfredson & Gottfredson, 2002; Ringwalt et al., 2011). When SEL programs *are* being used in schools, they are most often selected, implemented, and monitored in haphazard and uncoordinated ways that may thwart their success (Elias et al., 2015). In order to adopt, implement, and sustain SEL programs in schools, an infrastructure is needed to support the delivery of SEL programs (Fagan, Hawkins, & Shapiro, 2015; Shapiro, 2015). There is need for the systematic assessment of risk and protective factors to guide the implementation of effective prevention programs (Hawkins et al., 2015).

Risk and protective factors are terms used to characterize the predictors of MEB problems. Risk factors are characteristics or circumstances that make problems more likely to emerge, whereas protective factors are characteristics or resources that reduce the impact of risk, making it less likely that an individual will ultimately develop a MEB problem (Coie et al., 1993). The practices used to assess risk and protective factors are different than those used to determine the nature of an existing problem (Haggerty & Shapiro, 2013). Most child-serving systems have tools and processes for determining whether an individual case meets the criteria for an existing problem, in order to justify, inform, or monitor a remediation technique that could lessen the problem. On the other hand, many child-serving systems lack tools and processes for proactively identifying *threats* to well-being that justify, inform, or monitor a *prevention* technique that could avoid a MEB problem altogether.

To address this gap in tools and processes, many schools are developing their own internal infrastructures for prevention that use multi-tiered approaches (Maras, Thompson, Lewis, Thornburg, & Hawks, 2015). These tiered systems for early screening, prevention, and intervention, which became fundable through the *Individuals with Disabilities Education Act of 2004*, attempt to ensure each child receives, at minimum,

the type of prevention that is needed for college and career readiness, given constrained resources and classroom time. Ideally, these systems triage students based on the child's unique profile of risk and protection and their response to the least invasive preventive intervention. For example, a student might receive a *universal* intervention like the Promoting Alternative Thinking Strategies (PATHS; Kusche & Greenberg, 1994) curriculum, delivered to all students through a series of classroom-based lessons. Some students may also receive a *selective* intervention, directed only to a group of students determined to be at risk for a problem or who fail to respond to a high-quality, well-implemented universal intervention. Selective preventive interventions like the "Body Project" (Stice, Trost, & Chase, 2003) are often termed "Tier 2 interventions." Finally, some students may receive a Tier-3 (*indicated*) intervention if they are already experiencing early symptoms or signs of the problem, yet do not meet full diagnostic criteria for the condition (e.g., the "Blues Program"; Rohde, Stice, Shaw, & Brière, 2014). Determining which level of service is most appropriate is dependent on the availability of practical and psychometrically sound assessment tools that can screen a student population for predictors of MEB problems and facilitate decision making regarding the level of social emotional instruction that an individual child needs.

Guidance for how to screen a student population for predictors of MEB problems emerges from two distinct fields of practice: public health and school psychology. The World Health Organization published the Wilson-Jungner criteria in 1968, providing principles for screening that are still considered the "gold standard" in contemporary public health practice (Andermann, Blancquaert, Beauchamp, & Déry, 2008). More recently, these guidelines were adapted to be appropriate for screening risk and protective factors for the purpose of preventing MEB problems (O'Connell et al., 2009). These 10 prevention screening practice principles are paraphrased as follows:

1. The MEB problem should be considered a serious public health threat.
2. The relationship between the risk/ protective factors and the MEB problem should be clear.
3. Effective interventions to change the risk/protective factors should be identified and produce better outcomes than treatment after onset
4. Facilities or settings for screening and preventive intervention should be available.

5. Determination of level of risk/protective factors should require screening and assessment.
6. Screening tools or processes should be validated with acceptable accuracy compared with formal assessments.
7. Screening approaches and guidelines should be acceptable to the population and not lead to stigmatization.
8. Agreed upon guidelines should exist for whom to refer for assessment, prevention services, or treatment.
9. The cost of finding a case should be affordable, cost-effective, and reimbursable.
10. Case finding should be a continuous process.

The 1968 screening principles and their 2009 adaptation reflect a tension between the compelling and simple idea of early detection and intervention, and the actual challenge of wide-scale screening as a practice that results in an intervention for those with previously undetected needs, and avoids harm to those who are not in need (Wilson & Jungner, 1968). Although “universal screening to identify students at risk for school failure or psychological or behavioral problems is increasingly recognized as an important professional practice” (O’Connell et al., 2009, p. 229), screening for the prevention of MEB problems has not yet been widely adopted (Romer & McIntosh, 2005). Innovation is needed to establish screening tools and procedures that can balance Principle 9 (be quickly and easily performed) and Principle 6 (reasonable accuracy). Some instruments are available for collecting information about the risk and protective factors of students in school, such as the Communities That Care Youth Survey (Arthur, Hawkins, Pollard, Catalano, & Baglioni, 2002), the Youth Risk Behavior Surveillance Survey (Centers for Disease Control and Prevention, 2015), and the California Healthy Kids Survey (California Department of Education, 2015), but these instruments all use anonymous self-report forms that are generally analyzed at the population level. Approaches to collecting universal information about risk and protective factors in lower grades (in which children may not have self-reporting capacity), and in ways that can inform decision making about individual students, has been less of a focus in the public health literature.

The field of school psychology has contributed practice standards to guide universal screening in schools and child-care/after-school settings, particularly regarding their appropriateness for intended use, technical adequacy, and their usability in schools and after-school contexts

(Glover & Albers, 2007; Shapiro, Accomazzo, Claassen, & Fleming, 2015). General guidelines are also articulated in the *Standards for Educational and Psychological Testing* (Joint Committee on Standards for Educational and Psychological Testing, 2014). As clarified in the 2014 revision, validity refers to “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (Joint Committee on Standards for Educational and Psychological Testing, 2014, p. 11). A major source of evidence of validity comes from relations between test scores and variables external to the test. Concurrent criterion studies examine the relationship between a test score and a criterion contemporaneously, which are useful for determining the accuracy of screeners when the passage of time has not interfered with the presence of the underlying construct. Predictive criterion studies examine the relationship between a test score and a temporally delayed criterion, which is useful for determining the utility of a screener theorized to foreshadow the likelihood of an outcome over time. Both of these study designs are imperative for understanding whether a screening instrument does what it purports to do when used in multi-tiered SEL delivery systems in schools.

Common metrics for determining criterion validity include determinations of sensitivity, specificity, positive predictive value, and negative predictive value (Glover & Albers, 2007). Sensitivity is a determination of the extent to which a screening instrument correctly identifies those who are actually at risk. Specificity is a determination of the extent to which a screening instrument correctly identifies those who are not actually at risk. Positive predictive value (PPV) is a determination of the proportion of students who are *correctly* identified as at risk by a screening instrument out of all students who are identified as *at risk*, whereas negative predictive value (NPV) is a determination of the proportion of students who are *correctly* identified as *not at risk* by a screening instrument out of all of the students who are identified as not at risk.

In both the academic and behavioral realms, schools often use screening instruments in multi-tiered instructional models that have untested or inadequate performance against these metrics. Although it is desirable to have all indicators (sensitivity, specificity, PPV, and NPV) above 75% (Gredler, 2000; Kingslake, 1983), this is far more often achieved with narrow or targeted diagnostic screeners once a problem exists than with broad-based screening focused on risk and protective factors of individuals. Broad-based screeners tend to favor either sensitivity or specificity. Therefore, school practitioners face the decision as to whether their screening protocols will be over-inclusive, with higher



sensitivities triggering many false alarms, or over-exclusive, with higher specificities missing some students that may benefit from more intensive instruction (Levitt, Saka, Romanelli, & Hoagwood, 2007). For example, studies have revealed the DIBELS (Good & Kaminski, 2002) and the TPRI (Foorman et al., 1998) to have higher sensitivities and lower specificities (see Table 1). By using these popular screening instruments, schools are implicitly deciding that they would rather over-identify students for reading support than miss students who could benefit from more intensive reading instruction.

Schools may have different preferences for balancing sensitivity and specificity in the behavioral realm, in which substantial concerns persist with over-identification. These include (a) the potential to do harm through labeling by nature of the pervasive stigma surrounding mental illness, and (b) the barriers to a school's capacity to conduct an in-depth assessment to confirm case identification and offer more intensive interventions. When broad screeners for mental, emotional, and behavioral problems were reviewed by Levitt and colleagues in 2007 (Levitt et al., 2007), only four broad instruments (completed by adult informants) were identified with published studies of sensitivity and specificity (see Table 1). Unlike what can be observed among popular reading screeners, there is not a consistent pattern of over-identification. The Pediatric Symptom Checklist (PSC; Jellinek et al., 1988) and the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992) favor sensitivity (over-identification), whereas the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) and the Child Behavior Checklist (CBCL; Achenbach, 1991) favor specificity (under-identification).

An updated review by Jenkins and colleagues in 2014 (Jenkins et al., 2014) reiterated findings for the SDQ, and reported that sensitivity and specificity information on emerging screeners such as the Social Skills Improvement System (Elliott & Gresham, 2008) and the Systematic Screening for Behavior Disorders (Walker & Severson, 1992) was not available (see Table 1). Jenkins added information about the Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007), a newly published brief version of the BASC-2, which favored specificity. Given the general technical shortcomings of broad-based behavioral health screening tools to consistently be both sensitive and specific (275%), Glascoe (2005) advised practitioners who need to minimize excessive referrals to consider as an acceptable standard for screening instruments sensitivities of 50% and specificities of 80%, and PPV rates in the 30%–50% range. The BESS is the only screener known to meet Glascoe (2005) standards.

All of the universal screeners examined focused on students' problem behaviors; no reviewed instruments focused primarily on the risk or protective factors of individual children. Two promising instruments may have been overlooked by these reviews, likely for the following reasons. The Resiliency Scales for Children and Adolescents (RSCA; Prince-Embury, 2006) only has published RSCA sensitivity and specificity analyses that use gender and parent education level in addition to scale scores to determine correct classification rates (Prince-Embury, 2010), rendering the rates hard to interpret relative to other instruments or standards. The Social-Emotional Assets and Resilience Scales – Teacher Short Form (Merrell, 2011) manual does not report sensitivities or specificities, and Nese and colleagues (2012) called for accuracy analyses to be conducted on this tool as an important next step.

Thus far, we have discussed the *accuracy* of screening instruments intended to guide decisions related to the behavioral health of children in school. Of course, sensitivity and specificity are not the only instrument features in need of delicate balance. Universal prevention screening in schools needs to be practical for schools to implement and be capable of leading to service allocation decisions that improve the wellbeing and achievement of children. Both the public health and school psychology fields converge in their expectations that universal screening must be accurate and produce valuable/ actionable information, but also should not undermine teacher or parent discretion, be too time consuming, stigmatize students, or have embedded biases against groups of students (O'Connell et al., 2009).

The Devereux Student Strengths Assessment Mini (DESSA-Mini; Naglieri, LeBuffe, & Shapiro, 2011/2014) was designed to overcome practical obstacles to universal prevention screening. The DESSA-Mini is entirely strength-based, assessing only protective factors rather than risk factors or negative behaviors. Emerging evidence suggests that strength-based instruments may be as accurate as risk or problem-oriented instruments for identifying a case, without stigmatizing (or otherwise harming) individuals in a population (Dowdy, Furlong, Eklund, Saeki, & Ritchey, 2010; LeBuffe & Shapiro, 2004; Shochet, Dadds, Ham, & Montague, 2006). Furthermore, although many screeners depend on youth or parent reports so as to not overburden staff, universal participation presents an ongoing challenge for the purpose of screening. Other screeners rely on teacher reports, but require responses to 25 or more behavior rating scale items, spending a minimum of 5 min (and sometimes far more time) per child (Kamphaus et al., 2007). The DESSA-Mini contains only eight items, alleging a 1-min teacher completion rate per

child. Of course, this begs the question as to whether an entirely strength-based, eight-item screening instrument can achieve technical accuracy in routine practice.

In an initial attempt to answer this question, Naglieri, LeBuffe, and Shapiro (2011) calculated sensitivity, specificity, PPV, and NPV rates of the DESSA-Mini screener using the full 72-item DESSA (LeBuffe, Shapiro, & Naglieri, 2009/2014) as the criterion (a concurrent criterion study). Naglieri et al. (2011) reports sensitivities of .77 to .81 and specificities of .83 to .85. These rates not only exceed Glascoe (2005) standards for choosing among available instruments in ways that minimize excessive referrals, but also exceed the 75% ideal across screening instruments for diverse purposes. A strength of the Naglieri study was that it was conducted on a large ( $N = 1,234$ ) sample, representative of the diversity of children across the United States. The limitation of the Naglieri study was that the scores used to examine the relationships between the DESSA-Mini and the DESSA were obtained from a single administration of a single form. The full DESSA was administered during the standardization process that was used to develop the DESSA norms. Then, the item-level responses for the items that make up the DESSA-Mini were extracted from the full DESSA. The extracted DESSA-Mini items were then scored according to the DESSA-Mini manual, whereas the full DESSA was scored in accordance to the DESSA manual. Thus, Naglieri reports sensitivities and specificities of an extracted DESSA-Mini relative to the DESSA from which the items came. Although this presents strong evidence for the validity of using the DESSA-Mini as a universal prevention screener, the current article seeks to replicate the Naglieri method with a different sample of DESSAs, this time generated through routine practice; (b) conduct a concurrent criterion study in which the DESSA-Mini and full DESSA scores were collected independently within routine practice; and (c) conduct a predictive criterion study to determine the extent to which the DESSA-Mini administered early in the school year predicts which students have serious disciplinary infractions by the end of the academic year.

## **Method**

This study was conducted as part of a districtwide implementation of a new SEL initiative designed to promote students' social-emotional competence. The large, urban school district, situated in the northeastern United States, serves approximately 17,500 students across 23 schools. Students in this district are predominantly of Hispanic/Latino ethnicity (65%) and eligible for free or reduced priced lunch (80%) according to

district enrollment figures for the 2011–2012 academic school year (Pennsylvania Department of Education, 2012).

## **Design**

The SEL initiative intended to expose approximately 12,000 youth in district elementary and middle schools to preventive interventions across 3 years. During the 2011–2012 academic year, teachers of children in prekindergarten through second grade, across all 15 elementary schools, implemented the PATHS curriculum. Elementary teachers of children in third through fifth grade delivered no formal SEL instruction. In all four middle schools across the district, teachers of children in sixth through eighth grades delivered the Second Step curriculum (Committee for Children, 2008). The DESSA-Mini Form 1 was used to screen social-emotional competence through teacher ratings of all students in kindergarten through eighth grade near the start of the school year (October). To facilitate the study of concurrent validity, teachers also completed a DESSA on a random sample of approximately five children per classroom. District administrative data were used to determine serious disciplinary infractions during the academic year. Study protocols were reviewed and approved by the Institutional Review Board at the Devereux Foundation.

## **Sample**

Complete DESSA-Minis were obtained on 9,248 students, and complete DESSAs were obtained on 1,960 students. The current analyses used a subsample of the participants who had complete DESSA-mini and DESSA scores ( $n = 1,875$ ), comprised of students identified as 47% female and 65% Hispanic/Latino, 17% Black/African American, 14% White/European American, and 4% other races (i.e., multi-race, Asian/Pacific Islander American, Native American; see Table 2). Students from kindergarten through Grade 8 were included, with ages at the end of the school year ranging from 6 to 16 years ( $M = 9.7$ ). Race and gender were determined through school administrative records, whereas grade was collected from teachers at the time of DESSA rating.

## **Measures**

**Devereux Student Strengths Assessment (DESSA).** The DESSA (LeBuffe et al., 2009/ 2014) is a 72-item standardized, norm-referenced, strength-based behavior rating scale that assesses the social-emotional competence of children in kindergarten through eighth grade. The DESSA yields an overall total score called the Social-Emotional Composite (SEC)

as well as scores across eight domains of social-emotional competence, including Self-Awareness, Social Awareness, Self-Management, Goal-Directed Behavior, Relationship Skills, Personal Responsibility, Decision Making, and Optimistic Thinking (Smith, Shapiro, Sperry, & LeBuffe, 2014). The DESSA takes 10 min to administer and can be completed by parents or caregivers, teachers, out-of-school time program staff, staff at child-serving organizations, and other important adults in the child's life (Naglieri, LeBuffe, & Shapiro, 2013). The rater reads the stem "During the past four weeks, how often did the child . . ." and then rates each item on a 5-point Likert scale ranging from 0 to 4 (*never* = 0, *rarely* = 1, *occasionally* = 2, *frequently* = 3, *very frequently* = 4). Items are summed to raw scores, which are converted to *T* scores ( $M = 50$ ,  $SD = 10$ ), with high scores (*T* scores of 60 and above) indicating *strengths*, *T* scores between 41 and 59 (inclusive) representing *typical* scores, and *T* scores of 40 and below representing a *need for instruction*. The DESSA was designed to guide social-emotional instruction and interventions and for measuring progress and outcomes (Simmons, Shapiro, Accomazzo, & Manthey, 2016).

The DESSA was nationally standardized on a sample of 2,494 children and youth that closely approximated the kindergarten through eighth grade population of the United States at the time on the basis of age, gender, geographic region of residence, race, ethnicity, and socioeconomic status (LeBuffe et al., 2009/2014). The adequacy of the norms have been independently reviewed and determined to be representative, recent, and sufficiently large (Merrell & Gueldner, 2010). The DESSA has excellent internal reliability, with SEC alpha coefficients of .98 for parent raters and .99 for teacher/staff raters (LeBuffe et al., 2009/2014). Test–retest reliability of the SEC (the consistency of scores obtained for the same child when ratings were completed between 4 to 8 days apart) is .90 for parents and .94 for teachers. An examination of the absolute value of the test–retest difference for both parent and teacher raters on the SEC was less than one *T* score point (parent raters = 0.8; teacher raters = 0.6). Interrater reliability of the SEC (the consistency of scores obtained for a child by two different raters) is .78 for parent raters and .80 for teacher raters. Rater pairs differed, on average, by less than half a *T* score point (parent raters = 0.5; teacher raters = 0.4). Taken together, these results provide evidence that the DESSA is reliable for assessing children's social-emotional competence.

The DESSA has also demonstrated initial evidence of criterion and construct validity (LeBuffe et al., 2009/2014; Shapiro & LeBuffe, 2006). DESSA scores indicate very large differences ( $d = 1.39$ ) between samples

of typically developing students and those receiving special education services under the “seriously emotionally disturbed” (SED) classification. In addition, DESSA SEC scores, generated by teachers, co-vary with scores from the BASC-2 Behavioral Symptoms Index ( $r = -.72$ ) and Adaptive Skills Scale ( $r = .92$ ; Nickerson & Fishman, 2009). The DESSA has also been examined for racial and ethnic differences by rater, and found no interpretable differences for most comparisons. The exception was a small difference between how teachers rated the strengths of White students compared with Hispanic/Latino(a) students ( $d = 0.26$ ). A study presented in the test manual indicates that social-emotional competence, as measured by the DESSA, reduces negative outcomes in the context of risk, as one would expect from a scale designed to measure protective factors (LeBuffe et al., 2009/2014).

#### **Devereux Student Strengths Assessment Mini (DESSA-Mini).**

The DESSA-Mini (Naglieri et al., 2011/2014) is a brief, eight-item version of the DESSA that provides a snapshot of a student’s social-emotional competence. The DESSA-Mini was designed to be used for universal screening of social-emotional competence and ongoing progress monitoring. It can be completed in 1 min by teachers, out-of-school time staff, and staff at other child-serving organizations. The DESSA-Mini was developed by selecting items from the larger 72-item DESSA that most strongly correlated with the overall SEC. The DESSA-Mini has four parallel forms that can be used in rotation. Each yields a score referred to as the Social-Emotional Total (SET). Like the DESSA, results on the DESSA-Mini are provided as *T* scores, with high scores indicating *strengths*.

The DESSA-Mini norms were developed by utilizing the DESSA standardization sample of 1,250 children and youth rated by teachers and program staff. The DESSA-Mini has good internal reliability, with SET alpha coefficients exceeding .90 across each of the four forms (Naglieri et al., 2011/2014). The alternate form reliability of the DESSA-Mini is excellent, with correlation coefficient meeting or exceeding .90 across all form comparisons. Similarly, item means and standard deviations across the four forms (when rating the same children) are found to be very similar, with means ranging between 50.5 and 50.7, indicating their general interchangeability. Test–retest reliability of the four DESSA-Mini forms range between .88 and .94. An examination of the absolute value of the test–retest differences on the four DESSA-Mini forms ranged from 0 to 1.3 *T* score points. Interrater reliability of the DESSA-Mini forms ranged from .70 to .81 on the four forms. Rater pairs differed, on average, by less than one *T* score point across the forms, ranging from 0 to 0.6.

Taken together, these results provide evidence that the four DESSA-Mini forms are reliable for assessing children's social-emotional competence. The DESSA-Mini has demonstrated evidence of criterion and construct validity (Naglieri et al., 2011/2014). DESSA-Mini SET scores strongly correlate ( $r = .95-.96$ ) with DESSA SEC scores. Consistent classification of students in the *need for instruction* range between the DESSA-Mini and the DESSA is 94.5% to 95.3% across the four forms and were similar across children of different racial and ethnic groups (Naglieri et al., 2011/2014). Very large differences between the mean scores of typically developing students and those receiving special education services under the SED classification were found on all four DESSA-Mini forms ( $d$  ranged from 1.17 to 1.24).

**Disciplinary infractions.** In this district, disciplinary infractions are characterized into levels determined by the place and frequency of occurrence and the disruptive effect of the behavior on the learning environment. Level 1 infractions include a variety of behaviors that cause classroom disruption, independently managed by classroom teachers. Level 2 infractions are minor offenses reported to school administrators who respond in accordance to school-specific norms and guidelines. Level 3 infractions are more significant; responses to Level 3 infractions are consistent throughout the school district. Level 4 offenses generally result in suspension or expulsion. Level 3 and Level 4 disciplinary infraction records are tracked consistently in school administrative data. The prevalence of Level 3 and Level 4 infractions in the analysis sample is identical to the prevalence of Level 3 and Level 4 infractions among all kindergarten through eighth grade students across the district during the 2011–2012 academic year. Within the analysis sample, 10.0% ( $n = 188$ ) of students had Level 3 infraction records and 1.1% ( $n = 20$ ) had Level 4 infraction records. Although 93.1% ( $n = 175$ ) of students who had Level 3 infraction records did not have Level 4 infraction records, 65% ( $n = 13$ ) of students who had Level 4 infraction records had Level 3 infraction records. For the purpose of this analysis, a dichotomous variable was used to represent whether or not any serious (Levels 3 or 4) infraction record was present. This variable indicates that 10.4% ( $n = 195$ ) of students had any record of serious infraction by the end of the school year.

## **Analysis Plan**

We conducted several criterion relationship studies to explore the test accuracy of the DESSA-Mini: (a) a replication criterion study, a concurrent criterion study, and (c) a predictive criterion study.

**Replication criterion study.** Replication studies are essential for determining the generalizability of an instrument's validity evidence across contexts. We replicated the Naglieri et al. (2011) criterion study to examine the relationship between a test score and a criterion contemporaneously. We extracted DESSA-Mini items from the full DESSA, and then scored the Extracted-Mini and the DESSA in accordance to instructions contained within the respective DESSA-Mini and DESSA test manuals. The scores on the Extracted-Mini were used to determine the accuracy of the Extracted-Mini for identifying students in need of SEL instruction relative to determinations made based upon a full DESSA. Observed differences between our results and those of Naglieri may point to differences between samples, some of which may be related to differences between a researcher-derived standardization sample and a sample derived from routine practice.

Our analytic procedure consisted of five steps. First, we examined internal reliability. Generally, researchers have used Nunnally's (1978) .70 cut-off to claim acceptable reliability. However, Nunnally cautioned that in "applied settings where important decisions are made" internal reliability should be 2.90 (p. 245). Second, we compared the means and standard deviations between the Extracted-Mini SET scores and the DESSA SEC scores, and calculated the effect size of the difference between the scores using *d*-ratios. The *d*-ratio is a measure of the size of the difference between the mean scores, expressed in standard deviation units (Cohen, 1988). Cohen's *d*-ratio is interpreted as negligible if less than .2, and otherwise interpreted as small (2.2), medium (2.5), or large (2.8). Third, we examined the percentage of individuals who are classified (need for instruction vs. typical/strength) consistently based on their Extracted-Mini and DESSA scores. Fourth, we calculated sensitivity, specificity, PPV, and NPV of the Extracted-Mini relative to the DESSA. Finally, we conducted a receiver operating characteristic (ROC) curve analysis to understand how well the Extracted-Mini discriminates between students with needs for instruction and those who do not (Fawcett, 2006). Discriminant ability is measured by the area under the curve (AUC), ranging from 0.5 (no discrimination) to 1.0 (perfect discrimination), in which 2.90 is considered excellent, .80 to .90 is good, .70 to .80 is fair, .60 to .70 is poor, and .50 to .60 is a fail (Tape, 2001).

**Concurrent criterion study.** We conducted another concurrent criterion study in which the screener and the full assessment are collected independently. The DESSA-Mini was completed first by all teachers. The full DESSA was part of a battery of assessments that were completed in a counterbalanced order after the DESSA-Mini was completed. We



examined DESSA scores for mean differences by counterbalanced order, and no differences were found. The same five analytic steps conducted for the replication criterion study were repeated for the concurrent criterion study to assess the accuracy of the DESSA-Mini in relation to the DESSA. Observed differences between the results of our replication criterion study and this concurrent criterion study may be explained by variability introduced through test–retest scenarios (that are embedded in recommended use of the DESSA Comprehensive System).

**Predictive criterion study.** We conducted a predictive criterion study to examine the relationship between a test score and a temporally delayed criterion, which can be particularly useful for determining the utility of a screener theorized to foreshadow the likelihood of a future outcome. The DESSA-Mini, administered near the beginning of the school year, was used to predict which students have serious disciplinary infractions by the end of the academic year. We conducted a multilevel logistic regression (three levels to account for non-independence of observations within students, teachers, and schools) to test whether or not DESSA-mini SET scores predict students' disciplinary infraction records. Odds ratios were used to assess the likelihood of having a serious disciplinary infraction by the end of the year for those students who were identified near the beginning of the year as having a need for instruction, relative to those who were not identified as having a need for instruction.

## Results

### Replication Criterion Study

The Extracted-Mini showed excellent internal reliability at each grade ( $\alpha = .92$  to  $.95$ ; see Table 3). This compares favorably with the reliability coefficients of the Extracted-Mini derived from the standardization sample (.90 to .94) reported by Naglieri et al. (2011), and exceeds the stricter cutoff of .90 that Nunnally (1978) proposed for instruments that are used by professionals to make decisions.

The mean and standard deviations of the Extracted-Mini SET scores were higher than the DESSA SEC scores at every grade level (see Table 4), but these differences were generally too small to be interpretable ( $d < .20$ ). One exception was the difference between the Extracted-Mini SET and the DESSA SEC for kindergarten students, which narrowly met the criteria for a small difference in size ( $d = .21$ ). These findings from routine practice are generally consistent with the lack of differences between the Extracted-Mini and the DESSA in the standardization sample reported by Naglieri et al. (2011), which were all negligible.

Prediction consistency was high across all grades (See Table 5). In

this routine practice sample, 85% to 95% of students, across grades, were consistently identified by the Extracted-Mini and DESSA as either having a Need for Instruction or as not having a Need for Instruction. The median agreement consistency across grades of 94% in this routine practice sample is slightly lower than the agreement consistency (median of 96.5% across grades) reported by Naglieri et al. (2011), but provides evidence of high prediction consistency nonetheless.

The Extracted-Mini showed high sensitivity and specificity. Sensitivity, or the proportion of students with an actual Need for Instruction (DESSA <40) from among those who screened as having a Need for Instruction (Extracted-Mini <40), was 78%. This indicates that the Extracted-Mini correctly identifies most students with a need for SEL instruction. Specificity, or the proportion of students without an actual Need for Instruction (DESSA >40) from among those who screened as not having a Need for Instruction (Extracted-Mini >40), was 99%. This indicates that the Extracted-Mini correctly identifies nearly all students without a unique need for accelerated SEL instruction. PPV, or the proportion of students correctly screened as having a confirmed Need for Instruction (Extracted-Mini <40 and DESSA <40) from among all those who screened as having a Need for Instruction (Extracted-Mini <40), was 97%. NPV, or the proportion of students correctly screened as not having a confirmed Need for Instruction (Extracted-Mini >40 and DESSA >40) from among all those who screened as not having a Need for Instruction (Extracted-Mini >40) was 92%. Thus, the Extracted-Mini derived from this routine practice sample slightly privileges specificity preventing over-identifying students with needs, but exceeds ideals on both dimensions. The ROC analysis revealed an AUC of .88, indicating a good overall balance between sensitivity and specificity.

These findings are generally consistent with the sensitivity, specificity, PPV, and NPV of the Extracted-Mini generated from the standardization sample and reported by Naglieri et al. (2011). Although the sensitivity of the Extracted-Mini is the same in the routine practice sample as in the standardization sample, the specificity in the routine practice sample is notably higher (.99 relative to .85). The AUC of this routine practice sample (.88) is the same as the AUC of the standardization sample.

### **Concurrent Criterion Study**

When the DESSA-Mini was collected *independently* in routine practice, it continued to show excellent internal reliability ( $\alpha = .94$  to  $.96$ ), comparable with the Extracted-Mini ( $\alpha = .92$  to  $.95$ ; see Table 3). Similar to the findings reported above, the means and standard deviations of the

DESSA-Mini SET scores were higher than the DESSA SEC scores (see Table 4). As expected, the differences were larger between the total scores on the DESSA-Mini and the DESSA ( $d$ -ratio range = .13 to .48) than the differences were between the total scores on the Extracted-Mini and the DESSA ( $d$  = .01 to .21). Differences between the DESSA-Mini and the DESSA continue to be interpreted as negligible to small.

When the DESSA-Mini was collected independently in routine practice, prediction consistency continued to be high across all grades (see Table 5). Of the 1,875 students screened and assessed, 1,633 students (87%) were consistently identified by the DESSA-Mini and DESSA as either having a Need for Instruction or as not having a Need for Instruction. This classification rate, produced through a test retest scenario, is nearly as consistent as was found through extracting scores from a single administration (94%).

When the DESSA-Mini was collected independently in routine practice, it continued to meet Glascoe (2005) practice standards for sensitivity and specificity (see Table 6). Sensitivity was 62% and specificity was 98%. The PPV was 92%, and the NPV was 86%. Therefore although the DESSA-Mini had less impressive sensitivities and specificities when collected independently, it not only continued to meet Glascoe practice standards but also still met the 75% ideal across three of four of these indices, and only narrowly missed this ideal in sensitivity. The AUC of .79 indicates a fair balance between sensitivity and specificity. When collected independently in routine practice, the DESSA-Mini is again found to privilege specificity preventing the over-identification of students with needs for accelerated SEL instruction.

### **Predictive Criterion Study**

The DESSA-mini SET score in October significantly predicted the likelihood of a student having a serious disciplinary infraction by the end of the academic year. Students who were identified as having a Need for SEL Instruction in October were 4.5 times ( $d$  = .83) more likely to have a record of serious infraction by the end of the academic year compared with those who were not identified as having a Need for Instruction in October ( $p$  < .001; see Table 7).

### **Discussion**

Screening tools are necessary for the systematic assessment of risk and protective factors to guide the implementation of SEL programs in schools. Determining which level of service is most appropriate for an individual student is dependent on the availability of practical and psychometrically sound assessment tools that can screen a student

population for predictors of mental, emotional, and behavioral problems, and facilitate individual-level decision making. The DESSA-Mini was designed to overcome practical obstacles to universal prevention screening. This article sought to determine whether an entirely strength-based, eight-item screening instrument could achieve technical accuracy in routine practice. Across the replicated and new concurrent criterion studies conducted, excellent internal reliability was repeatedly found exceeding Nunnally's (1978) criteria of .90. We also found only "negligible" to "small" differences between scores on the DESSA-Mini screener and the DESSA full assessment. Classification consistency between the DESSA-Mini and the DESSA was also high across studies, averaging 87% to 94% in routine practice. All sensitivity and specificity estimates exceeded Glascoe (2005) practice standards and only narrowly missed the 75% ideal across sensitivity, specificity, PPV, and NPV indices. These findings compare quite favorably with studies of other instruments used in schools to screen entire student populations, and are consistent with a practice preference not to over-identify students for accelerated preventative MEB interventions. The field has called for predictive criterion studies using outcomes of interest to schools, like Office Disciplinary Referrals (King & Reschly, 2014). We learned through this predictive criterion study that students who were identified as having a Need for SEL Instruction on the DESSA-Mini in October were 4.5 times more likely to have a serious disciplinary infraction by the end of the academic year relative to those who were not identified.

This series of studies replicates and expands upon previous work in several ways. First, these studies sought to isolate the difference between a researcher-driven sample and a practice-driven sample. It appears that a practice-driven sample yielded very similar results to the research-driven sample, even though the samples themselves had some noteworthy differences. In other words, there were differences between the sample used by Naglieri et al. (2011) and the one used in this article, other than one being research-driven and one being practice-driven. The Naglieri sample included teacher and staff raters, whereas the current sample included only teacher raters. Furthermore, the Naglieri sample was representative of the U.S. population by region, whereas the current sample was derived from only one large school district in the Northeast. The samples had different race and ethnicity compositions, and the current sample had a more consistent age distribution than the Naglieri sample.

Second, these studies sought to isolate the difference between a sample derived from a single administration (in which Mini scores were extracted from the full assessment) and a sample derived from a process

more similar to routine practice (in which the Mini is followed by a full assessment sequentially). The DESSA-Mini Form used in this study has a test–retest reliability of .94, in which *T* scores are separated by an average of .6 points. Thus, we expected (and found) that the sequenced administration of the same items would yield some reduction in consistent classification. The multiple administrations performed in this study, however, were separated by less time than may be typical in practice. The passage of more time could mean less consistent classification, particularly when the passage of time also introduces a change of rater, which happens because of high turnover of educators.

Third, these studies sought to explore (a) the relationship between a test score and a criterion contemporaneously, such that the passage of time had not interfered with the presence of the underlying construct, as well as (b) the relationship between a test score and a temporarily delayed criterion, such that the utility of a screener theorized to foreshadow a future event could be tested. It should be noted that many of the students that participated in this study received a prevention program during the year, which could change the relationship between the screen and the outcome. This study used disciplinary infractions as a criterion. Although disciplinary records are important indicators of school outcomes, they have a potential to carry bias (e.g., race, gender; Skiba, Michael, Nardo, & Peterson, 2002) and may be an imperfect proxy for a child’s problem behavior. In hopes of reducing the impact of teacher-level bias, this study used only the most serious disciplinary infractions with consistent district-level standards. Future studies should conduct predictive criterion studies with additional outcome indicators. The 2014 revision of the *Standards for Educational and Psychological Testing* (Joint Committee on Standards for Educational and Psychological Testing, 2014) suggest that when a test is used to classify students to different levels of service, additional support for the validity of the classification decision is desirable. This evidence could be derived by studying how the various groups benefit from classification decisions. Thus, we suggest follow-up research to examine how DESSA-Mini scores at the beginning of the school year determine growth of protective factors in response to a preventive intervention.

In summary, we found across studies that the DESSA-Mini compares well with standards of the field in terms of internal reliabilities, effect size differences, consistent classification rates, sensitivities, specificities, positive predictive values, negative predictive values, and discrimination rates. No matter how these indices were examined, this 1-min screener does not seem to compromise accuracy in order to be a practical tool for the systematic assessment of protective factors to guide

the implementation of effective prevention programs in schools.

## References

- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4 –18 and 1991 Profile*. Burlington, VT: University of Vermont.
- Andermann, A., Blancquaert, I., Beauchamp, S., & Déry, V. (2008). Revisiting Wilson and Jungner in the genomic age: A review of screening criteria over the past 40 years. *Bulletin of the World Health Organization*, *86*, 317–319. <http://dx.doi.org/10.2471/BLT.07.050112>
- Arthur, M. W., Hawkins, J. D., Pollard, J. A., Catalano, R. F., & Baglioni, A. J., Jr. (2002). Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors: The Communities That Care Youth Survey. *Evaluation Review*, *26*, 575– 601.
- California Department of Education. (2015). *California Healthy Kids Survey*. Retrieved from [https:// chks.wested.org/](https://chks.wested.org/)
- Centers for Disease Control and Prevention. (2015). *Youth Risk Behavior Survey*. Retrieved from [http:// www.cdc.gov/yrb](http://www.cdc.gov/yrb)s
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., . . . Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*, *48*, 1013–1022. <http://dx.doi.org/10.1037/0003-066X.48.10.1013>
- Committee for Children. (2008). *Second step: Student success through prevention for middle school*. Seattle, WA: Author.
- Doll, B. (1996). Prevalence of psychiatric disorders in children and youth: An agenda for advocacy by school psychology. *School Psychology Quarterly*, *11*, 20 – 46. <http://dx.doi.org/10.1037/h0088919>
- Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children’s social and emotional competence: A randomized trial of the preschool “PATHS” curriculum. *The Journal of Primary Prevention*, *28*, 67–91. <http://dx.doi.org/10.1007/s10935-007-0081-0>
- Dowdy, E., Furlong, M., Eklund, K., Saeki, E., & Ritchey, K. (2010). Screening for mental health and wellness: Current school-based practices and emerging possibilities. In B. Doll, W. Phofl, & J. Yoon (Eds.), *Handbook of youth prevention science* (pp. 70 –95). New York, NY: Routledge.
- Doyle, A., Ostrander, R., Skare, S., Crosby, R. D., & August, G. J. (1997). Convergent and criterion related validity of the Behavior

- Assessment System for Children-Parent Rating Scale. *Journal of Clinical Child Psychology*, 26, 276–284.  
[http://dx.doi.org/10.1207/s15374424jccp2603\\_6](http://dx.doi.org/10.1207/s15374424jccp2603_6)
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405–432.  
<http://dx.doi.org/10.1111/j.1467-8624.2010.01564.x>
- Elias, M. J., Leverett, L., Duffell, J. C., Humphrey, N., Stepney, C., & Ferrito, J. (2015). Integrating SEL with related prevention and youth development approaches. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 33–49). New York, NY: Guilford Press.
- Elliott, S. N., & Gresham, F. K. (2008). *SSIS performance screening guide*. Minneapolis, MN: Pearson Assessments
- Fagan, A. A., Hawkins, J. D., & Shapiro, V. B. (2015). Taking SEL to scale in schools: The role of community coalitions. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 468–481). New York, NY: Guilford Press.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861–874.  
<http://dx.doi.org/10.1016/j.patrec.2005.10.010>
- Foorman, B. R., Fletcher, J. M., Francis, D. J., Carlson, C. D., Chen, D., Mouzaki, A., . . . Taylor, R. H. (1998). *Technical report for the Texas Primary Reading Inventory* (1998 ed.). Houston, TX: Center for Academic and Reading Skills and the University of Houston
- Glascoc, F. P. (2005). Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews*, 11, 173–179.  
<http://dx.doi.org/10.1002/mrdd.20068>
- Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, 45, 117–135. <http://dx.doi.org/10.1016/j.jsp.2006.05.005>
- Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills, administration and, scoring guide* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Child Psychology & Psychiatry & Allied*



- Disciplines*, 38, 581–586. <http://dx.doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40, 1337–1345. <http://dx.doi.org/10.1097/00004583-200111000-00015>
- Gottfredson, D. C., & Gottfredson, G. D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency*, 39, 3–35. <http://dx.doi.org/10.1177/002242780203900101>
- Gredler, G. R. (2000). Early childhood screening for developmental and educational problems. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (3rd ed., pp. 399 – 411). Needham Heights, MA: Allyn & Bacon.
- Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist*, 58, 466 – 474. <http://dx.doi.org/10.1037/0003-066X.58.6-7.466>
- Haggerty, K. P., & Shapiro, V. B. (2013). Science-based prevention through communities that care: A model of social work practice for public health. *Social Work in Public Health*, 28, 349 –365. <http://dx.doi.org/10.1080/19371918.2013.774812> (Erratum, 2014, *Social Work in Public Health*, 29, p. 87).
- Hawkins, J. D., Jenson, J. M., Catalano, R. F., Fraser, M. W., Botvin, G. J., Shapiro, V. B., . . . the Coalition for Behavioral Health. (2015). *Unleashing the power of prevention* [Discussion paper]. Washington, DC: Institute of Medicine and National Research Council.
- Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review*, 32, 541–556.
- Individuals with Disabilities Education Act of 2004, 20 U.S.C., §§ 1400 – 1482 (2004).
- Jellinek, M. S., Murphy, J. M., Robinson, J., Feins, A., Lamb, S., & Fenton, T. (1988). Pediatric Symptom Checklist: Screening school-age children for psychosocial dysfunction. *The Journal of Pediatrics*, 112, 201–209. [http://dx.doi.org/10.1016/S0022-3476\(88\)80056-8](http://dx.doi.org/10.1016/S0022-3476(88)80056-8)

- Jenkins, L. N., Demaray, M. K., Wren, N. S., Secord, S. M., Lyell, K. M., Magers, A. M., . . . Tennant, J. (2014). A critical review of five commonly used social-emotional and behavioral screeners for elementary or secondary schools. *Contemporary School Psychology, 18*, 241–254. <http://dx.doi.org/10.1007/s40688-014-0026-6>
- Jensen, P. S., Watanabe, H. K., Richters, J. E., Roper, M., Hibbs, E. D., Salzberg, A. D., & Liu, S. (1996). Scales, diagnoses, and child psychopathology: II. Comparing the CBCL and the DISC against external validators. *Journal of Abnormal Child Psychology, 24*, 151–168. <http://dx.doi.org/10.1007/BF01441482>
- Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.
- Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 Behavioral and Emotional Screening System manual*. Bloomington, MN: Pearson.
- Kamphaus, R. W., Thorpe, J. S., Winsor, A. P., Kroncke, A. P., Dowdy, E. T., & VanDeventer, M. C. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and Psychological Measurement, 67*, 342–356. <http://dx.doi.org/10.1177/00131644070670021001>
- King, K. R., & Reschly, A. L. (2014). A comparison of screening instruments: Predictive validity of the BESS and the BSC. *Journal of Psychoeducational Assessment, 32*, 687–698. <http://dx.doi.org/10.1177/0734282914531714>
- Kingslake, B. (1983). The predictive (in) accuracy of on-entry to school screening procedures when used to anticipate learning difficulties. *Special Education: Forward Trends, 10*, 23–26. <http://dx.doi.org/10.1111/j.1467-8578.1983.tb00184.x>
- Kusche, C. A., & Greenberg, M. T. (1994). *The PATHS (Promoting Alternative Thinking Strategies) curriculum*. South Deerfield, MA: ChanningBete.
- LeBuffe, P. A., & Shapiro, V. B. (2004). Lending “strength” to the assessment of preschool social-emotional health. *California School Psychologist, 9*, 51–61. <http://dx.doi.org/10.1007/BF03340907>
- LeBuffe, P. A., Shapiro, V. B., & Naglieri, J. A. (2014). *The Devereux Student Strengths Assessment (DESSA): Assessment, technical manual, and user’s guide*. Charlotte, NC: Apperson, Inc. (Original work published 2009)

- 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. <http://dx.doi.org/10.1016/j.jsp.2006.11.005>
- Maras, M. A., Thompson, A. M., Lewis, C., Thornburg, K., & Hawks, J. (2015). Developing a tiered response model for social-emotional learning through interdisciplinary collaboration. *Journal of Educational & Psychological Consultation, 25*, 2–3, 198–223. <http://dx.doi.org/10.1080/10474412.2014.929954>
- Merrell, K. W. (2011). *Social and Emotional Assets and Resilience Scales (SEARS)*. Lutz, FL: Psychological Assessment Resources.
- Merrell, K. W., & Gueldner, B. A. (2010). *Social and emotional learning in the classroom: Promoting mental health and academic success*. New York, NY: Guilford Press.
- Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. B. (2011). Universal screening for social emotional competencies: A study of the reliability and validity of the DESSA-mini. *Psychology in the Schools, 48*, 660–671. <http://dx.doi.org/10.1002/pits.20586>
- Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. B. (2013). Assessment of social-emotional competencies related to resilience. In S. Goldstein & R. Brooks (Eds.), *Handbook of resilience in children* (pp. 261–272). New York, NY: Kluwer/Academic Press. [http://dx.doi.org/10.1007/978-1-4614-3661-4\\_15](http://dx.doi.org/10.1007/978-1-4614-3661-4_15)
- Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. B. (2014). *The Devereux Student Strengths Assessment – Mini (DESSA-Mini): Assessment, technical manual, and user’s guide*. Charlotte, NC: Apperson. (Original work published 2011)
- Nese, R. N., Doerner, E., Romer, N., Kaye, N. C., Merrell, K. W., & Tom, K. M. (2012). Social Emotional Assets and Resilience Scales: Development of a strength-based short-form behavior rating scale system. *Journal for Educational Research Online, 4*, 124.
- Nickerson, A. B., & Fishman, C. (2009). Convergent and divergent validity of the Devereux Student Strengths Assessment. *School Psychology Quarterly, 24*, 48–59. <http://dx.doi.org/10.1037/a0015147>
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- O’Connell, M., Boat, T., & Warner, K. E. (Eds.). (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington, DC: National Academies Press.

- Payton, J. W., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). *Positive impact of social and emotional learning for kindergarten to eighth grade students: Findings from three scientific reviews* [Technical report]. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.
- Pennsylvania Department of Education. (2012). *Enrollment public schools 2011–2012*. Retrieved from [http://www.portal.state.pa.us/portal/server.pt/community/enrollment/7407/public\\_school\\_enrollment\\_reports/62054](http://www.portal.state.pa.us/portal/server.pt/community/enrollment/7407/public_school_enrollment_reports/62054)
- Prince-Embury, S. (2006). *Resiliency Scales for Children and Adolescents: A profile of personal strengths*. San Antonio, TX: Harcourt Assessments.
- Prince-Embury, S. (2010). Assessment for integrated screening and prevention using the Resiliency Scales for Children and Adolescents. In B. Doll., W. Phofl, & J. Yoon (Eds.), *Handbook of youth prevention science* (pp. 141–162). New York, NY: Routledge.
- Reynolds, C. R., & Kamphaus, R. W. (1992). *BASC: Behavior Assessment System for Children: Manual*. Circle Pines, MN: American Guidance Service.
- Ringwalt, C., Vincus, A. A., Hanley, S., Ennett, S. T., Bowling, J. M., & Haws, S. (2011). The prevalence of evidence-based drug use prevention curricula in U.S. middle schools in 2008. *Prevention Science, 12*, 63–69. <http://dx.doi.org/10.1007/s11121-010-0184-3>
- Rohde, P., Stice, E., Shaw, H., & Brière, F. N. (2014). Indicated cognitive behavioral group depression prevention compared to bibliotherapy and brochure control: Acute effects of an effectiveness trial with adolescents. *Journal of Consulting and Clinical Psychology, 82*, 65–74. <http://dx.doi.org/10.1037/a0034640>
- Romer, D., & McIntosh, M. (2005). The roles and perspectives of school mental health professionals in promoting adolescent mental health. In D. L. Evans, E. B. Foa, R. E. Gur, H. Hendin, C. P. O'Brien, M. E. P. Seligman, & B. T. Walsh (Eds.), *Treating and preventing adolescent mental health disorders: What we know and what we don't know* (pp. 597–615). New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/9780195173642.003.0032>
- Shapiro, V. B. (2015). Resilience: Have we not gone far enough? A response to Larry Davis. *Social Work Research, 39*, 7–10. <http://dx.doi.org/10.1093/swr/svv001>
- Shapiro, V. B., Accomazzo, S., Claassen, J., & Fleming, J. L. (2015).

- Multi-method social emotional assessment in out of school time settings: Choices, challenges, and lessons learned. *Journal of Youth Development: Bridging Research and Practice*, 10, 32–45.
- Shapiro, V. B., & LeBuffe, P. A. (2006). Using protective factors in practice: Lessons learned about resilience from a study of children aged five to thirteen. *Annals of the New York Academy of Sciences*, 1094, 350–353. <http://dx.doi.org/10.1196/annals.1376.048>
- Shochet, I. M., Dadds, M. R., Ham, D., & Montague, R. (2006). School connectedness is an underemphasized parameter in adolescent mental health: Results of a community prediction study. *Journal of Clinical Child and Adolescent Psychology*, 35, 170–179. [http://dx.doi.org/10.1207/s15374424jccp3502\\_1](http://dx.doi.org/10.1207/s15374424jccp3502_1)
- Simmons, C. A., Shapiro, V. B., Accomazzo, S., & Manthey, T. J. (2016). Strengths-based social work: A meta-theory to guide social work research and practice. In N. Coady & P. Lehmann (Eds.), *Theoretical perspectives for direct social work practice* (3rd ed., pp. 131–154). New York, NY: Springer.
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*, 34, 317–342. <http://dx.doi.org/10.1023/A:1021320817372>
- Smith, G. T., Shapiro, V. B., Sperry, R. W., & LeBuffe, P. A. (2014). A strengths-based approach to supervised visitation in child welfare. *Child Care in Practice*, 20, 98–119. <http://dx.doi.org/10.1080/13575279.2013.847056>
- Stice, E., Trost, A., & Chase, A. (2003). Healthy weight control and dissonance-based eating disorder prevention programs: Results from a controlled trial. *International Journal of Eating Disorders*, 33, 10–21. <http://dx.doi.org/10.1002/eat.10109>
- Tape, T. G. (2001). Interpretation of diagnostic tests. *Annals of Internal Medicine*, 135, 72. <http://dx.doi.org/10.7326/0003-4819-135-1-200107030-00043>
- Walker, H. M., & Severson, H. (1992). *Systematic screening for behavior disorders: Technical manual*. Longmont, CO: Sopris West.
- Wilson, J. M. G., & Jungner, G. (1968). *Principles and practice of screening for disease*. Geneva, Switzerland: World Health Organization. Retrieved from <http://www.who.int/bulletin/volumes/86/4/07-050112bp.pdf?ua=1>