Mapping the Research Base for Universal Behavior Screeners

Katie Scarlett Lane Pelton[1], Kathleen Lynne Lane[2], Wendy Peia Oakes[3], Mark M. Buckman[2],

David James Royer[4], Rebecca Lee Sherod[2]

[1] University of Connecticut – Neag School of Education,

[2] University of Kansas

[3] Arizona State University

[4] University of Louisville

Author Note

Katie Scarlett Lane Pelton 0000-0002-7640-6651 (Katie.Lane@uconn.edu)

Kathleen Lynne Lane 0000-0001-6364-838X (Kathleen.Lane@ku.edu)

Wendy Peia Oakes 0000-0002-3533-8293 (Wendy.Oakes@asu.edu)

Mark M. Buckman 0000-0001-9332-0940 (Buckman@ku.edu)

David James Royer 0000-0003-2882-1049 (David.Royer@louisville.edu)

Rebecca Lee Sherod 0000-0001-7486-9217 (RebeccaSherod@ku.edu)

Abstract

Universal behavior screening is used in schools worldwide to detect students with and at risk for behavioral challenges. A plethora of instruments is available for this purpose, though little metascience has been conducted to review and synthesize methods used to study these instruments in educational settings nor is there a comprehensive list of instruments to support educators in selecting an appropriate tool. We conducted this review to provide a rigorous – and accessible – overview of the research base for universal behavior screening instruments to facilitate educators' decision-making process when selecting a systematic screening tool for the students they serve and identify areas of further refinement for the research community. This scoping review includes an extensive list of behavior screening instruments, an examination of how these tools have been studied, and areas for future research. We identified 56 behavior screening instruments. The most common psychometric analyses included coefficient alpha for internal consistency, correlations between theoretically related variables, and confirmatory factor analysis. We discuss other methods currently employed as well as methods and complexities for consideration in future research.

Keywords: behavior screening, scoping review, instrument validation, psychometrics

Mapping the Research Base for Universal Behavior Screeners

Many school-age youth experience social, emotional, and behavioral health concerns at some point during their school years. Forness and colleagues (2012) compiled point prevalence estimates of students meeting criteria for any emotional or behavioral disorder (EBD) – internalizing (e.g., anxiety) or externalizing (e.g., aggression) – during their school years and conservatively estimated 12% with included studies ranging from 3.7% to 21.1%. Prevalence rates, particularly for internalizing disorders (e.g. anxiety, depression), have been on the rise in recent years among school-aged youths (Lebrun-Harris et al., 2022). For example, in the 2018-2019 National Survey of Children's Health, 13.2% of U.S. children ages 13-17 had a diagnosed mental or behavioral health condition (Health Resources and Services Administration, 2020). More recently, in the 2023 Youth Behavior Risk survey, 29% of high school students reported poor mental in the last 30 days (Centers for Disease Control and Prevention, 2024). Although internalizing disorders are far less disruptive than externalizing behaviors in the classroom environment, they can still have detrimental impacts on students' school experiences and may contribute to poor mental health (Weist et al., 2018). Collectively, estimates indicate many students need additional support at some point in their school career, though internalizing behaviors are often more challenging to recognize (Bradshaw et al., 2008).

Many schools adopted a systemic approach to examine overall levels of EBDs in school systems, identify students experiencing these challenges, and connect students with evidence-based strategies at the first sign of concern before the COVID-19 pandemic and these systems are now being recommended as a way to support students' social, emotional, and behavioral needs as part of pandemic recovery (Office of Special Education and Rehabilitative Services, 2022; Walker et al., 2014). Tiered systems such as Positive Behavioral Interventions and

Supports (PBIS; Sugai & Horner, 2002), Comprehensive, Integrated, Three-tiered (Ci3T) model of prevention (Lane et al., 2009), Multi-Tiered System of Supports (MTSS; McIntosh & Goodman, 2016), and Interconnected Systems Framework (ISF; Barrett et al., 2017) share a key element: universal screening. When implementing these systems as designed, schools screen all students in one or more domain: academic, behavioral, and social and emotional well-being. Just like vision or hearing screenings, academic, behavioral, and social-emotional screening involves brief assessments used with all students intended to measure the specific constructs of interest (e.g., externalizing or internalizing behaviors; Oakes et al., 2017). Educators can use systematic screening data to shape instructional practices at Tier 1 for all students which may include low-intensity strategies (e.g., precorrection, instructional choice) to maximize engagement and limit challenging behavior (Korpershoek et al., 2016; Ma et al., 2022). Additionally, screening within a tiered system can create a clear and equitable path for connecting students to Tier 2 and 3 interventions when Tier 1 efforts – even when implemented as planned – are not sufficient to meet students' needs (Lane, Menzies et al., 2020). Tier 2 and 3 interventions may include validated interventions in the academic, behavioral, mental health, and/or social-emotional learning domains (e.g., Cipriano et al., 2023; Cohen et al., 2024, Murano et al., 2020; Sabey et al., 2017). As schools prepare to implement these systems, they must select a screening tool.

Several scholars have published recommendations to support education leaders in selecting a screening assessment. Glover and Albers (2007) recommended first evaluating the appropriateness for intended use which includes making sure the (a) tool fits the need, (b) construct of interest aligns with their purpose, (c) tool has been validated for a similar population, and (d) instrument has theoretical and empirical support. Second, schools or districts evaluate technical adequacy of instruments meeting the first set of standards by ensuring the (a)

normative sample included students with similar characteristics to the student population they are serving, (b) measure demonstrated adequate reliability properties (e.g., internal consistency, test-retest stability, inter-rater reliability), and (c) studies have established adequate validity evidence (e.g., criterion, construct, content, prediction). Third, the tool must be feasible given cost, social validity, school systems or infrastructure, available accommodations, and applicability of results. Oakes and colleagues (2017) provided a similar framework for selecting a behavior screening tool: determine constructs of interest (e.g., internalizing and externalizing), narrow down potential instruments based on feasibility (e.g., cost, time), then evaluate psychometric properties. American Education Research Association (AERA), American Psychological Association (APA), and National Council for Measurement in Education (NCME; 2014) provide further guidance regarding instrument development and criteria for selecting an instrument in *The Standards for Educational and Psychological Testing.* The National Center on Intensive Interventions (NCII, 2022) provided specific criteria for educational screening. These collective recommendations offer a roadmap for selecting a screening tool, though this roadmap may be challenging for schools and districts to implement as information on systematic screening tools has developed substantially in the last decade (Lane et al., 2012), leaving a widely dispersed set of information to review.

Aim and Scope

Initially, we planned to compile psychometric information on behavior screening instruments to assist practitioners in meeting established guidelines for selecting a screening instrument. As we prepared to conduct this review, we continually located more relevant instruments, though no comprehensive list of these instruments nor a consolidated source of information across instruments existed. We found several prior reviews with narrower scopes

such as a single instrument (Kersten et al., 2015; Lambert, Sointu, et al., 2018; Marzocchi et al., 2004), a smaller selection of instruments (Caselman & Self, 2008; Jenkins et al., 2014) or a specific age range (Whalen et al., 2017). These reviews provide important depth to their respective areas, yet these were not designed to detail the full scope of available instruments to support practitioners in the early stages of instrument selection.

As such, we shifted our focus towards conducting a thorough scoping review to provide practitioners and researchers with a comprehensive yet accessible set of information regarding (a) behavior screening tools available to detect a range of behavior patterns of interest to aid in instrument selection and (b) summarize analyses used to study these tools to inform future research. We offer this rigorous and accessible scoping review of behavior screening tools to support schools in selecting instruments, identify current research practices, and highlight areas for refinement in future research (Boveda et al., 2023; Munn et al., 2018). We address two research questions. First, what systematic behavior screening tools are available for use in schools with students in grades preK-12? Second, how have researchers studied the psychometric properties of these instruments?

Method

We conducted a multi-phase search initially guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (Page et al., 2021). After our focus shifted towards a scoping review of universal behavior screening research, to maintain the methodological rigor and extend the potential reach to guide decisions regarding systematic screening tool selection (Boveda et al., 2023), we used the PRISMA extension for scoping review to update our process and reporting (Tricco et al., 2018). Our search included three phases with reliability assessed in each phase: (1) an electronic database search using Boolean

operators, (2) hand searches through identified journals, and (3) a second electronic database

search including additional instruments identified in the first two phases (see Figure 1). See

LDBase.org for a summary of findings from each included study.

Inclusion Criteria

The first author trained all authors how to apply the following inclusion criteria to ensure

accuracy of the article selection process. They used a PowerPoint and provided each co-author a

draft of the inclusion criteria below to reference throughout the coding process. All authors

independently rated three articles to practice applying criteria, with 100% agreement.

Published in English

The study may take place in any country though the article must be published in English

so the research team can determine if the study meets inclusion criteria. Studies published in

other languages (e.g., Bacanli & Erdoğan, 2003) were not included, a noted limitation.

Published in a Peer-Reviewed Journal

The study must be published in a peer-reviewed journal. We did not include dissertations

and theses as we could not determine the rigor of review, given the variability of different

university requirements. Yet, many methodologically sound theses and dissertations are later

published in journals and could then be included. We did not have any non-examples based on

search parameters (see Table S1 and Table S2).

Administered in a PreK-12 Educational Setting

The study must have been conducted in a school environment or educational context such

as residential, self-contained, private, university lab, or public school. We excluded studies in

clinical settings and community-based programs (e.g., Lambert et al., 2015; Walrath et al., 2004)

as our research question focused on providing guidance for selecting screening tools for use in

preK-12 educational settings.

Administered Universally

The study must include a behavior screening tool administered to, or offered pending consent and assent, all students in a set age or grade range. Many studies did not provide detailed information on their sampling procedures, so we could not create a more precise decision rule beyond universal administration. Studies including only subset of students pre-identified by another characteristic (e.g., disability status, teacher referral) were excluded (e.g., Hysing et al., 2007; Treyvaud, 2014). Furthermore, studies including a subset of students selected randomly or for non-specified reasons were not included (e.g., Shojaei et al., 2008).

Behavior Screener

Because little meta-science has been conducted in this field, there is no broadly accepted operational definition for a behavior screener. We focused on identifying studies that included a brief, universal assessment for characteristic behavior patterns of major childhood behavior disorders (i.e., internalizing and/or externalizing; Achenbach, 1991). Screening instruments are typically brief, though they vary in length based on the number and depth of constructs assessed and stage in instrument development as they are often pruned over time (Oakes et al., 2017). We excluded studies using only extended rating scales intended for diagnostic or special education eligibility purposes such as the Teacher Report Form (TRF; Achenbach & Rescorla, 2001) and Social Skills Improvement System Rating Scales (SSiS-RS; Gresham & Elliot, 2008) as they are not brief screening instruments intended for universal implementation (e.g., Flanagan et al., 1996; Gresham et al., 2010).

Complete Instrument

The study must use a complete screening tool or an entire subscale (e.g., Stage II

measures of the Systematic Screening for Behavior Disorders; SSBD; Walker et al., 2014) as designed or intentionally study a revised version of a tool. If the instrument was edited or modified (e.g., modified response anchors, unintentionally omitted items) and the changes are not the primary focus or conclusions are related to the original instrument, we excluded the study (e.g., King & Reschly, 2014; Moulton & Young, 2021) as seemingly minor instrument changes can impact findings (Goodman et al., 2007).

Psychometric Analysis

The study must present at least one psychometric analysis of a screening tool to indicate the study evaluated the instrument itself in some way. The study could focus on reliability or reliability evidence of any kind (AERA et al., 2014). We did not include practice-guides (e.g., Lane et al., 2011) and application studies (e.g. screening tool used as a pre- and post-test for an intervention study) as they do not provide information on how instruments are studied (e.g., Kamps et al., 2003).

Primary Database Search

Prior to developing Boolean search terms, we conducted informal searches through existing literature (e.g., Caselman & Self, 2008; Jenkins et al., 2014; Lane et al., 2012), practice briefs (e.g., Lane, 2019), and professional learning content about systematic behavior screening (e.g., ci3t.org/screening; pbis.org) to create a preliminary list of instruments. We determined a preliminary list was necessary to limit the initial search given the variety of language used to describe behavior screening instruments and relevant analyses. Using approximately 30 different sources, we compiled an initial list of screening tools and behavior rating scales often used as comparison for screening tools. Next, we contacted senior scientists and researchers working with districts to implement systematic screening to ensure all screening tools they were aware of

were included on this list. Although this was not a comprehensive search, we expected to find additional screening tools in articles referencing other established measures in the introduction or as a comparison tool. Any additional screening instruments found in the primary and hand searches were included in a secondary electronic search, conducted approximately one year following the primary search.

We conducted the electronic database search in ERIC and APA PsycINFO databases – two databases with substantial representation of relevant educational research that have been used in other scoping reviews related to measurement in tiered systems (e.g., Buckman et al., 2021). The primary search included terms to specify the study occurred in a school setting, addressed psychometric properties, and included at least one behavior rating tool (Table S1). We used check boxes to limit results to articles published in English by peer reviewed journals. We conducted the primary search through the University of Kansas library on December 22, 2021. A second author ran the same search terms and returned the same articles with 100% reliability. The primary search yielded 2,163 articles.

Titles and Abstracts Coding

The first author reviewed all 2,163 abstracts to determine if they met inclusion criteria, erroring on the side of inclusion when there was no clear reason for exclusion. Two other authors served as secondary coders and assessed 25% of abstracts ($n = 569$) for reliability. Primary and secondary coders reached 94.71% interrater agreement (IRA; Cohen, 1960) on the binary decision for inclusion in the next step with Cohen's kappa ($\kappa = 0.84$, 95% CI $= [0.79, 0.90]$) in the near perfect agreement range per Landis and Koch (1977).

Full Read Coding

From the abstracts, we identified 362 articles for a preliminary full text read. Co-authors

completed an average of 10 full article reads per week for 8-10 weeks, enabling the first author

to conduct weekly reliability checks to prevent drift between raters. Each author read

approximately one fourth of articles in total to determine if the article was eligible for inclusion.

The first author checked reliability on a minimum of 25% of articles coded by each co-author.

Overall agreement between the first author and reliability raters was 90.29% (κ = 0.81, 95% CI =

[0.69, 0.92], indicating very high agreement). Three articles had substantial ambiguity regarding

procedures and/or measures. Authors met to reach consensus on these articles which led to

clarification on how to determine if an instrument qualified as a screener for EBD. We decided

to include all brief instruments implemented universally and measuring at least one behavior

characteristic of EBD. From this preliminary read, we identified 122 articles for inclusion.

Journal Hand Search

We identified 15 journals with three or more published articles included in the primary

search to search by hand online. The journals are listed here in alphabetical order: *Assessment*,

Assessment for Effective Intervention, *Behavioral Disorders*, *European Journal of Psychological*

Assessment, *International Journal of School & Educational Psychology*, *Journal of Child and*

Family Studies, *Journal of Emotional and Behavioral Disorders*, *Journal of Psychopathology*

and Behavioral Assessment, *Journal of Psychoeducational Assessment*, *Journal of School*

Psychology, *Psychological Assessment*, *Remedial and Special Education*, *School Mental Health*,

School Psychology, and *School Psychology Review*. For these journals, we reviewed all issues

from date of the first included article (1987) or earliest issue available online (whichever was

later) through September 30, 2022.

The first author reviewed all 23,101 titles and a secondary coder assessed a minimum of

25% of issues (*n* = 622) from each hand-searched journal. Raters noted the number of articles in

each issue to ensure they checked the same titles and abstracts and reached 99.64% agreement

(range = 97.87 to 100%). Average agreement for number of articles to include from each issue

was 95.75% (range = 85.71 to 100%). We identified 251 articles for full read based on titles and

abstracts, of which 147 (58.57%) were previously read in full in the primary search. The first

author read all 104 unique articles and another author completed reliability coding of 25% of

articles (n = 26; IRA = 92.31%; κ = 0.84, 95% CI = [0.62, 1.05], indicating high agreement). We

identified 25 new studies for inclusion, of which five were published after we conducted the

primary search.

Secondary Database Search

As expected, we identified additional screening instruments from the primary and hand

searches. We added search terms for these instruments to the original terms from the primary

electronic database search and removed criterion measures (e.g., TRF, SSiS-RS) as they were

included only to assist in identifying additional screening tools (supplemental Table 2). Besides

screening tool specific search terms, we used the same university library, databases, and search

parameters as the primary search. We conducted the secondary search on January 2, 2023. Two

authors conducted the search separately and returned the same 1,816 articles with 100%

reliability, yielding 353 new articles, including articles published after the primary search.

Titles and Abstracts Coding

Following the same procedures as the primary search, the primary coder reviewed all

titles and abstracts except those already read in full (n = 1,491). Two other authors served as

reliability coders for 25% of abstracts (n = 373; IRA = 94.10%; κ = 0.74, 95% CI = [0.63, 0.84],

indicating substantial agreement).

Full Read Coding

From abstracts, we identified 459 articles for a preliminary read of which 125 were already included from the primary and hand searches. The first author read all new articles and two other authors scored at least 25% for reliability (n = 94), yielding 96.81% agreement (κ = 0.90, 95% CI = [0.79, 1.01], indicating very high agreement). From this preliminary read, we identified 33 new articles to include for information extraction.

Information Extraction

Finally, authors read all included articles (n = 180) in full and recorded key information, primarily from method and results sections, in a coding spreadsheet designed for systematic reviews available at ci3t.org. We recorded information on participant demographic information, school level, geographic location, locale (when available), screening informant(s), and all included instruments. Then we recorded psychometric analyses reported and a brief overview of findings. Another author checked coding for 25% of articles (n = 47) and found 98.16% average accuracy (range = 92.86 to 100%). See our coding sheet at LDBase.org for sample characteristics from all included articles.

<div align="center">

Results

</div>

We begin by providing an overview of included studies. Many studies involved multiple school levels (i.e., elementary, middle, high school), informants, geographic regions, and analyses. The findings presented are not mutually exclusive, so percentages in each category do not sum to 100%. Studies included elementary-aged students most often (n = 110; 61.11%), followed by middle school (n = 67; 37.22%), high school (n = 42; 23.33%), and preschool or early childhood (n = 19; 10.56%). Teachers served as informants most often (n = 117; 65.00%), followed by students (n = 68; 37.78%), and parents (n = 35; 19.44%). Of the included studies, 119 (66.11%) included samples from North American countries of which 116 (64.44%) were in

the United States, 35 (19.44%) from Europe, 18 (10.00%) from Asia, eight (4.44%) from

Australia, four (2.22%) from South America, two (1.11%) from Africa, and three (1.67%) with

location not explicitly reported. Our coding sheet with information on the samples from each

article and summaries of findings is available at LDBase.org, though not all articles reported the

same depth of demographic characteristics. Notably, describing demographic characteristics of

parent or teacher raters was an exception rather than the norm. The earliest included study was

published in 1987 and our searches concluded with articles published in 2022. We present the

cumulative total of studies examining psychometric properties of behavior screening instruments

administered universally in a celeration graph (Figure 2; Kennedy, 2005) which shows slow

growth in the body of literature for the first 20 years and greater increase in the last 15 years.

Identified Instruments

To address question one in assisting practitioners in identifying potential screening

instruments appropriate for the students they serve and their intended use, we created a list of all

behavior screening instruments located throughout the search. We identified 56 instruments

(Table 1) and listed articles included for each instrument in Supplemental Table 3. Wording for

instrument and subscale names varies greatly across instruments due to different foci (e.g.,

internalizing, externalizing) and approaches (i.e., deficit skills, strength-based). Instrument

length ranged from 3 to 66 items with nearly all instruments utilizing a Likert-type scale with 3

to 11 response options.

Reliability Analyses

The most reported analysis was estimates of internal consistency which were used in 149

articles (82.78%). Coefficient alpha was by far the most reported ($n = 110$; 61.11%) with more

authors including omega in recent years ($n = 21$; 11.67%) and few other metrics ($n = 18$;

10.00%). 12 articles (6.67%) included both alpha and omega estimates (Black et al., 2021; Español-Martín et al., 2021; Gillé et al., 2021; Gomez & Stavropoulos, 2020; Margherio et al., 2019; Pierce et al., 2016; Renshaw, 2019; Renshaw & Cook, 2019; Sharma et al., 2022; Volpe et al., 2021; von der Embse et al., 2017; Xu et al., 2021), five articles (2.78%) included alpha and another reliability estimate (Aitken et al., 2015; Garrido et al., 2020; Gustafsson et al., 2016; Kilgus et al., 2018; Naser, Brown, et al., 2018), and one article (0.56%) included alpha, omega, and Spearman-Brown estimates (Naser, Hitti, et al., 2018).

Several studies ($n = 43$; 23.89%) examined test-retest reliability or temporal stability of behavioral ratings from two weeks to over a year. Most test-retest reliability analyses involved correlations ($n = 39$; 21.67%) with relatively few other analyses methods employed: intraclass correlations (ICC; $n = 3$; 1.67%; Aanondson et al., 2020; Erdogan & Ozturk, 2011; Isolan et al., 2011), kappa coefficients ($n = 2$; 1.11%; Feil et al., 1995; Pagano et al., 2000), percent at risk ($n = 1$; 0.56%), and ANOVA ($n = 1$; 0.56%). Fewer studies evaluated inter-rater reliability ($n = 31$; 17.22%). Correlations were the most common method for assessing inter-rater reliability ($n = 28$; 15.56%) followed by intraclass correlations ($n = 4$; 2.22%; Caldarella et al., 2008; Downs et al., 2012; Kilgus et al., 2015; Mieloo et al., 2014) and kappa coefficients ($n = 3$; 1.67%; Feil et al., 1995; Kilgus et al., 2015; Margherio et al., 2019). ICCs can also be used to estimate the degree to which students within a class had similar scores by estimating the proportion of score variance due to the cluster, or class (McCoach & Cintron, 2022). Very few studies reported these ICCs ($n = 8$; 4.44%; Eklund et al., 2017; Kilgus et al., 2018; Kilpatrick et al., 2018; Liu et al., 2020; Splett et al., 2017; Splett et al., 2018; von der Embse et al., 2016; Wiesner & Schanding, 2013).

Structural Analyses

Many studies evaluated the internal structure of instruments with factor analysis.

Confirmatory factor analysis (CFA), often reported in relation to construct validity, was the third

most common analysis overall ($n = 77$; 42.78%). Fewer studies reported exploratory factor

analysis (EFA; $n = 45$; 25.00%) though more than half of those reporting EFA also reported CFA

($n = 25$; 13.89%). Multiple group confirmatory factor analysis (MGCFA) or measurement

invariance was even less common ($n = 34$; 18.89%), though all studies evaluated the fit of the

measurement model with CFA first. MGCFA has become more common with 31 of the 34

studies published in the last decade. Structure of the Strengths and Difficulties Questionnaire

(SDQ; Goodman, 1997), including many translations, was frequently studied accounting for over

a third of CFA ($n = 26$), EFA ($n = 15$), and MGCFA ($n = 12$) studies. Relatively few studies

examined the structure of the instrument using item response theory (IRT) analyses. Only 12

(6.67%) articles reported some form of IRT analysis of which eight (4.44%) included differential

item functioning analyses (Barbarin et al., 2020; Bøe et al., 2016; Bosik et al., 2022; Deighton et

al., 2013; Harrel-Williams et al., 2015; Kim & Kamphaus, 2018; Lambert et al., 2014;

Schatschneider et al., 2014).

Relational Analyses

Researchers established validity evidence by comparing screening scores to other

theoretically related variables such as other screeners, behavior rating scales, or educational

outcomes (e.g., attendance, office discipline referrals [ODRs], academic outcomes). Many

studies reported correlation coefficients ($n = 84$; 46.67%), though correlations are often a

prerequisite for more advanced analyses. Conditional probabilities ($n = 47$; 26.11%) paired with

receiver-operator characteristic curves ($n = 33$; 18.33%) were the next most common relational

analyses followed by regression ($n = 41$; 22.78%), and analysis of variance techniques (e.g.,

ANOVA, MANOVA; $n = 14$; 7.78%). Few studies reported other techniques such as kappa

coefficients ($n = 7$; 3.89%), t-tests ($n = 8$; 4.44%), structural equation modeling ($n = 6$; 3.33%) and graphical examination ($n = 2$; 1.11%).

Cluster Analyses

Seven articles (3.89%) presented clustering analyses to examine patterns of behavioral risk over time (Bauer, 2022). These studies included *k*-mean cluster analysis ($n = 1$; 0.56%; Dever et al., 2017), latent profile analysis ($n = 3$; 1.67%; Dowdy et al., 2014; Kilgus et al., 2015; Warmbold-Brann et al., 2017), latent class analysis ($n = 2$; 1.11%; Kilgus et al., 2019; King et al., 2016), and latent transition analysis ($n = 1$; 0.56%; Iaccarino et al., 2019).

Social Validity

Twelve articles (6.67%) included some analysis of social validity – the acceptability of goals, procedures, and outcomes of universal behavior screening (Wolf, 1978) – to give voice to those completing the screening tools. Researchers examined social validity via response rates (e.g., Lane et al., 2010) or instruments such as the Assessment Rating Profile-Revised (ARP-R; Daniels et al., 2017; Eckert et al., 1999), Screening Tool Rating Scale (STR; Lane & Oakes, 2010; Lane et al., 2014; Lane et al., 2015; Oakes et al., 2016), and Usage Rating Profile– Assessment (URP-A; Chafouleas et al., 2012; Hartman et al., 2017).

Discussion

Educators around the world use universal behavior screening to inform instructional experiences for all learners, connect students with validated interventions when the core curriculum available to all students, or Tier 1, is insufficient, and direct teacher efforts to respectfully maximize instruction (Gresham et al., 2013; Lane et al., 2022). Although a wealth of behavior screening tools exists, there was not an extensive list to assist educators beginning the instrument selection process and there was little metascience on the field at large. Our goal was

to provide a rigorous, accessible review of universal behavior screening instruments (Boveda et al., 2023). This scoping review resulted in an extensive list of universal behavior screeners and mapped the methods used to study these tools.

Identified Instruments

We found far more screening instruments than expected – 56 in total – though we suspect many tools measure slightly different constructs. For example, social-emotional and behavior screeners, especially those with a strength-based approach, often contain similar language, though the constructs are not necessarily interchangeable (Lane, Oakes, Monahan et al., 2023). Both are valuable tools for informing instruction and intervention efforts, though they do not have the same intended use in practice thus requiring different validity evidence to align with their respective uses (Kane, 2013). We encourage instrument developers to clearly define intended constructs, researchers to engage in critical conversations on how to clarify boundaries of these constructs, and educators to consider which constructs align best with their intended use when selecting an instrument.

In recent years, the number of universal screening studies has increased substantially. The bulk of the evidence base for universal behavior screening has been established at the elementary level suggesting there is ample room to study these tools in early childhood, middle school, and high school settings. Teachers were the most common informant in school-based screening studies which seems fitting when the purpose is to inform educational experiences and detect students in need of additional support at school specifically. Students were the second most common informant followed by parents. Additional information gleaned from student or familial raters brings additional and important perspectives to inform students' behavioral needs, especially beyond the school setting (De Los Reyes & Epkins, 2023). However, school system

must be prepared to respond swiftly and effectively when new information is gleaned from students and families given ethical and legal obligations as mandated reporters (Lane et al., 2021). Most studies in this sample included samples from the United States, though our requirement of being published in English certainly impacted the geographical distribution of studies, a noted limitation.

For educators interested in implementing universal screening, there is a wealth of literature available. This is encouraging given systematic screening tools are one potential path to equitably connecting students with supports aligned with their level of needs. Beyond this review, a wide range of practice guides are available on pbis.org and ci3t.org/screening for educators and technical assistance providers to not only assist with selecting screening instruments (e.g., Lane, 2019; Oakes, Buckman, et al., 2021) but how to set up structures to prepare to implement screening (e.g., Lane, Oakes, Menzies, et al., 2020; Oakes et al., 2022, Schonour et al., 2022) and use screening tools to shape instructional experiences for students (e.g., Lane, Powers, et al., 2020; Ma et al., 2022).

Analysis Methods

As we examine methods used to study these instruments, we consider the evolution of research methodology as well as the chronological entry-point of any given instrument over the 45-year range of included studies. We do not devalue a study using a technique considered dated by today's standards when it was cutting-edge or an expected analysis at time of publication. Similarly, we recognize each of these instruments was once, or may still be, in its infancy. We encourage further inquiry into tools with less published research or fewer modern analyses as they may be efficient and effective tools for educators. Even for established instruments, validation is an ongoing process as populations of interest, understanding of a construct, and

inferences made from the scores may change over time (McCoach et al., 2013). We encourage researchers to reevaluate the utility of each instrument regularly, with continued research and refinement viewed as a strength of the instrument rather than a limitation.

In terms of reliability evidence, we were encouraged to see researchers moving away from relying solely on coefficient alpha as the underlying assumptions are rarely met fully in practice (Cronbach & Shavelson, 2004; Sijtsma & Pfadt, 2021). Instead, researchers can report other estimates of reliability with more plausible assumptions – particularly regarding tau equivalence – in addition to coefficient alpha (McNeish, 2018; Revelle & Condon, 2019). Test-retest and inter-rater reliability studies are relatively common, though we encourage researchers to consider their underlying theory when conducting these studies. When examining test-retest reliability, behaviors may not be stable over extended periods (e.g., over one year), especially if the student is receiving intervention in response to a prior screening score. Similarly, we encourage researchers to consider how student behavior may differ across settings when selecting multiple raters for inter-rater reliability as true differences in behavior may impact these reliability estimates (Achenbach et al., 1987).

Confirmatory factor analysis was the second most reported analysis which is a well-known process for confirming observed data fits with the hypothesized structure of the instrument. Conversely, item response theory approaches to item evaluation are relatively under-utilized with behavior screeners. IRT allows for individuals' scores and screener items to be placed along the same continuum of latent trait ability to determine one's location and an item's discrimination (de Ayala, 2022). Thus, IRT may be particularly useful to identify highly informative items for brief screeners to ensure adequate information around specific ability levels, like those used as cut scores (Embretson, 1996). Computer adaptive testing, an advanced

application of IRT, could even be used to create hybrid screening and diagnostic behavioral assessments to more swiftly and precisely identify students' behavioral strengths and areas for support (Chang, 2004). Multiple group confirmatory factor analysis and differential item functioning analyses are becoming more common as methods to detect measurement bias (Berry, 2015). We anticipate these analyses will continue to become more commonplace as researchers work to support equitable educational practices and new criteria are established for screening instruments (e.g., NCII, 2022).

We grouped analyses relating screening instruments to other variables (e.g., predictive, discriminative, convergent, and concurrent validity) to align with more recent conceptualizations of validity evidence (AERA et al., 2014). Correlations were the most common analysis, which is expected when preceding other analyses but leaves much room for expansion when reported alone. We hope more researchers will use analysis methods which can appropriately model the known nested structure of the data with students nested within their classrooms (Huang, 2018). In the eight studies reporting ICC estimates for cluster-effects, ICCs ranged from .02 to .55, empirically confirming multilevel modeling would be useful to appropriately account for nesting (Geldhof et al., 2014). Overlooking cluster or rater effects can lead to underestimated standard errors and increase Type I error rate (McCoach & Cintron, 2022). There are several methods for correcting standard errors, though researchers in social sciences utilize hierarchical linear modeling most often (McNeish et al., 2017). Although these analyses are more complex to conduct and interpret, we believe collaboration with methodological experts may result in more rigorous and nuanced understanding of how behavior screening instruments function in schools (e.g., Lane, Oakes, Buckman, et al., 2023; Reinke et al., 2022).

Mixture modeling methods such as latent class analysis, latent profile analysis, and latent

transition analysis were also relatively uncommon as they appeared in only seven of the included articles. These techniques are known as person-centered approaches to summarizing data by grouping observations by similar response patterns (Schmiege et al., 2017). Mixture models can help researchers retroactively evaluate how the tool functioned for grouping students, which may be particularly useful when studying these tools in the context of tiered systems, though they should be used with caution for diagnostic purposes. Even when using the correct mixture model, individuals are not always classified into the correct group especially with small or unequal groups (Cintron et al., 2023). In the context of tiered systems, screeners are often used to identify approximately 15% of students with some additional need and connect them with Tier 2 interventions and connect approximately 5% of students with the most intensive needs to Tier 3 interventions (Gresham et al., 2013; Lane et al., 2009), although it is likely these percentages have increased in the aftermath of the COVID-19 pandemic (Weist et al., 2023). Given the unequal and progressively smaller group sizes, mixture modeling may be inaccurate for determining classification and needs to be used with caution.

Social validity was among the least common though longest-standing analyses identified in this review. Stemming from behavioral intervention research, social validity is a measure of the acceptability of goals, procedures, and outcomes as determined by invested parties (Wolf, 1978). For universal screening, social validity among informants is particularly important to understand barriers to implementation and ensure future participation in universal screening. We found several instruments available for assessing social validity for researchers interested in studying this further (e.g., ARP-R, Eckert et al., 1999; STR, Lane & Oakes, 2010; URP-A, Chafouleas et al., 2012). We encourage additional inquiry in this area to uncover barriers to universal screening implementation which may show the need for further research and

professional learning to address educator concerns.

Over the 45 years of school-based behavior screening research, researchers have utilized a broad range of methods. We are encouraged to see growth of literature aimed towards supporting educators and students in schools. In future articles, we recommend researchers provide demographic information on both informants and students when available and include detailed information on their procedures for data collection, especially if they limited their sample by some criteria. In terms of analyses, we suggest authors include two or more estimates of reliability by instrument subscale each time they evaluate an instrument. We also encourage researchers to report correlations between subscales and other measures used in the study prior to more complex analyses including methods which appropriately account for nesting of students within classroom. With ever-expanding analytic options, we encourage researchers to carefully consider which method best aligns with their research questions and available data and select the most appropriate and robust methods to evaluate behavior screening instruments.

Limitations

The primary limitations of this scoping review relate to our inclusion criteria. First, we specified the article must be published in a peer-reviewed journal. Studies with null or negative results may not have been included due to publication bias (Cook et al., 2017). We hope the open science movement, coupled with the understanding null effects also provide important contributions, will minimize the effects of this limitation in coming years. For logistical reasons, our team was limited to studies published in English which makes this scoping review more relevant for predominately English-speaking countries than other locales. We also limited our search to studies in which the screener was administered universally. Some studies used robust sampling procedures; however, many studies included little description of their sampling

procedures or utilized techniques that could introduce bias into the study. Often, we could not evaluate the strength of the sampling procedures, so we agreed upon the universally administered criterion to ensure relevance for educators seeking a universal screening tool. Open science practices are calling for increasingly clear methods for future replication and this limitation could be mitigated with those practices. The universal criterion also eliminated studies which examined screening for a subset of a school's student population (e.g., English language learners, children in foster care, refugees). Although outside the scope of this review, we recognize the need and merit in studying the functionality of these instruments with more specific populations and we strongly encourage practitioners to reference those studies when they have a particular population they are hoping to screen. Despite these limitations, we are hopeful this review will assist educators in locating screening tools and inform future research in universal behavior screening.

Summary

We conducted this review to create a rigorous and comprehensive overview of the field including an extensive list of behavior screening instruments for implementers and map how instruments have been studied to date. We located 56 instruments available for screening and 180 articles examining the psychometric properties of these tools meeting our criteria. The most common analyses examined internal consistency, correlations with related variables, and factor analysis. We encourage researchers investigating systematic screening to continue to evolve with research methodology to utilize the most appropriate and robust analytic techniques for their research questions. We hope in another five to 10 years the field will have continued the ongoing validation process for these instruments to ensure we are accurately and equitably screening for emotional and behavioral disorders in our school age population. In this meantime, we are

hopeful this systematic review will be useful to researchers as they conduct additional psychometric inquiry using current data analytic methodologies. In addition – and perhaps most importantly – we hope this review will be useful to educational leaders to inform their decision as to which screening tools to adopt to inform instruction within integrated systems to support educators in providing positive and productive learning environments for preK-12 students.

References

* = Articles included in the review; all included articles listed in Supplemental Table 3

*Aanondsen, C. M., Jozefiak, T., Heiling, K., & Rimehaug, T. (2020). Validation of the

   Strengths and Difficulties Self-Report in Norwegian Sign Language. *The Journal of Deaf*

   Studies and Deaf Education, 25(1), 91–104. https://doi.org/10.1093/deafed/enz026

Achenbach, T. M. (1991). *Manual for the child behavior checklist/4-18 and 1991 profile.*

   University of Vermont, Department of Psychiatry.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral

   and emotional problems: Implications of cross-informant correlations for situational

   specificity. *Psychological Bulletin, 101*(2), 213–232. https://doi.org/bqcsgb

Achenbach, T. M., McConaughy, S. H., Ivanova, M., & Rescorla, L. A. (2011). *Manual for the*

   ASEBA Brief Problem Monitor. University of Vermont, Research Center for Children,

   Youth, & Families.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms &*

   Profiles. University of Vermont, Research Center for Children, Youth, & Families.

*Aitken, M., Martinussen, R., Wolfe, R. G., & Tannock, R. (2015). Factor structure of the

   Strengths and Difficulties Questionnaire in a Canadian elementary school sample.

   Assessment for Effective Intervention, 40(3), 155–165. https://doi.org/mjk8

American Education Research Association, American Psychological Association, and National

   Council of Measurement in Education. (2014). *Standards for Educational and*

   Psychological Testing. AERA. https://www.aera.net/Publications/Books/Standards-for-

   Educational-Psychological-Testing-2014-Edition

*Arslan, G. (2019). Development and validation of the Youth Externalizing Behavior Screener: A brief and effective measure of behavioral problems. *International Journal of School & Educational Psychology, 7*(1), 64–74. https://doi.org/10.1080/21683603.2018.1466747

*Arslan, G. (2021). Measuring emotional problems in Turkish adolescents: Development and initial validation of the Youth Internalizing Behavior Screener. *International Journal of School & Educational Psychology, 9*(2), 198–207. https://doi.org/10.1080/21683603.2019.1700860

Bacanli, H., & Erdoğan, F. (2003). Adaptation of the Matson Evaluation of Social Skills with Youngsters (MESSY) to Turkish. *Kuram ve Uygulamada Eğitim Bilimleri*, *2*, 368–379. https://doi.org/ 10.5172/jamh.1.3.235

Barbarin O. (2004). *ABLE: Screening for problems of attention, aggression, behavior, language and emotional self-regulation in young children*. HPT1, LLC.

*Barbarin, O., Hitti, A., & Brown, J. (2020). Assessing the severity of concerns about preschool children's self-regulation of attention, behavior, and emotions using The ABLE universal screener: A Rasch analysis. *Journal of Emotional and Behavioral Disorders, 28*(3), 167–179. https://doi.org/10.1177/1063426619864932

Barrett, S., Eber, L., & Weist, M. (2017). *Advancing educational effectiveness: Interconnecting school mental health and school-wide positive behavior support*. https://www.pbis.org/resource/advancing-education-effectiveness-interconnecting-school-mental-health-and-school-wide-positive-behavior-support

Bauer, J. (2022). A primer to latent profile and latent class analysis. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for Researching Professional Learning and Development: Challenges, Applications and Empirical Illustrations* (pp. 243–268).

Springer.

Behar, L. (1977). The Preschool Behavior Questionnaire. *Journal of Abnormal Child
Psychology, 5*(3), 265–275. https://doi.org/10.1007/BF00913697

Berry, C. (2015). Differential validity and differential prediction of cognitive ability tests:
Understanding test bias in the employment context. *Annual Review of Organizational
Psychology and Organizational Behavior*, *2*, 435-463. https://doi.org/gckfm4

Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & Neer, S. M. (1997).
The Screen for Child Anxiety Related Emotional Disorders (SCARED). *Journal of the
American Academy of Child & Adolescent Psychiatry, 36* (4), 545-553.

*Black, L., Mansfield, R., & Panayiotou, M. (2021). Age appropriateness of the self-report
strengths and difficulties questionnaire. *Assessment, 28*(6), 1556–1569.
https://doi.org/10.1177/1073191120903382

*Bøe, T., Hysing, M., Skogen, J. C., & Breivik, K. (2016). The Strengths and Difficulties
Questionnaire (SDQ): Factor Structure and Gender Equivalence in Norwegian
Adolescents. *PLOS ONE, 11*(5), e0152202. https://doi.org/10.1371/journal.pone.0152202

*Bosik, S. W., Harris, B., Alibrahim, N., & Donovan, C. (2022). Construct validity of the teacher
version of the strengths and Difficulties Questionnaire in a sample of elementary-aged
children in the United States. *Contemporary School Psychology, 26*(4), 518–534.
https://doi.org/10.1007/s40688-021-00371-8

Boveda, M., Ford, K. S., Frankenberg, E., & López, F. (2023). Editorial vision 2022–2025.
Review of Educational Research, 93(5), 635–640. https://doi.org/gtgbqp

Bradshaw, C. P., Buckley, J. A., & Ialongo, N. S. (2008). School-based service utilization among

urban children with early onset educational and mental health problems: The squeaky

wheel phenomenon. *School Psychology Quarterly, 23*(2), 169–186. https://doi.org/cvjwz2

Buckman, M. M., Lane, K. L., Common, E. A., Royer, D. J., Oakes, W. P., Allen, G. E., Lane,

K. S., & Brunsting, N. (2021). Treatment integrity of primary (tier 1) prevention efforts

in tiered systems: Mapping the literature. *Education and Treatment of Children*.

https://doi.org/10.1007/s43494-021-00044-4

*Caldarella, P., Young, E. L., Richardson, M. J., Young, B. J., & Young, K. R. (2008).

Validation of the Systematic Screening for Behavior Disorders in middle and junior high

school. *Journal of Emotional and Behavioral Disorders, 16*(2), 105–117.

https://doi.org/10.1177/1063426607313121

Caselman, T. D., & Self, P. A. (2008). Assessment instruments for measuring young children's

social-emotional behavioral development. *Children & Schools*, *30*(2), 103–115.

https://doi.org/10.1093/cs/30.2.103

Centers for Disease Control and Prevention. (2024). *Youth Risk Behavior Survey Data Summary

& Trends Report: 2013–2023.* U.S. Department of Health and Human

Services. https://www.cdc.gov/yrbs/dstr/index.html

Chafouleas, S. M., Kilgus, S. P., Jaffery, R., Riley-Tillman, T. C., Welsh, M., & Christ, T. J.

(2013). Direct behavior rating as a school-based behavior screener for Elementary and

middle grades. *Journal of School Psychology, 51*(3), 367–385.

https://doi.org/10.1016/j.jsp.2013.04.002

Chafouleas, S., Miller, F., Briesch, A., Neugebauer, S., & Riley-Tillman, C. (2012). *Usage

Rating Profile–Assessment*. University of Connecticut.

Chang, H. H. (2004). Understanding computerized adaptive testing: From Robbins-Munro to

> Lord and beyond. In D. Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology*
>
> *for the Social Sciences* (pp. 117-133). New York: Sage.

Cintron, D. W., Loken, E., & McCoach, D. B. (2023). A cautionary note about having the right

> mixture model but classifying the wrong people. *Multivariate Behavioral Research,*
>
> *58*(4), 675–686. https://doi.org/10.1080/00273171.2022.2082913

Cipriano, C., Naples, L. H., Eveleigh, A., Cook, A., Funaro, M., Cassidy, C., McCarthy, M. F.,

> & Rappolt-Schlichtmann, G. (2023). A systematic review of student disability and race
>
> representation in universal school-based social and emotional learning interventions for
>
> elementary school students. *Review of Educational Research*, *93*(1), 73-102.
>
> https://doi.org/10.3102/00346543221094079

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological*

> *Measurement*, *20,* 37-46.

Cohen, K. A., Ito, S., Ahuvia, I. L., et al. (2024). Brief School-Based Interventions Targeting

> Student Mental Health or Well-Being: A Systematic Review and Meta-Analysis. *Clinical*
>
> *Child and Family Psychology Review, 27,* 732–806. https://doi.org/10.1007/s10567-024-
>
> 00487-2

Conduct Problems Prevention Research Group. (1995). *Teacher social competence scale*

> *technical report*. Fast Track Project.

Cook, B. G., & Therrien, W. J. (2017). Null effects and publication bias in special education

> research. *Behavioral Disorders, 42*(4), 149-158. https://doi.org/gc6zhx

*Cook, C. R., Rasetshwane, K. B., Truelson, E., Grant, S., Dart, E. H., Collins, T. A., & Sprague,

> J. (2011). Development and validation of the Student Internalizing Behavior Screener:

Examination of reliability, validity, and classification accuracy. *Assessment for Effective Intervention, 36*(2), 71–79. https://doi.org/10.1177/1534508410390486

Cook, C., Volpe, R., & Gresham, F. M. (2012). *Technical adequacy, classification accuracy, and social validity of the Student Externalizing Behavior Screener*. Unpublished manuscript.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391–418. https://doi.org/10.1177/0013164404266386

Cullinan D., & Epstein M. H. (2013). *Emotional and Behavioral Screener.* PRO-ED.

*Daniels, B., Volpe, R. J., Fabiano, G. A., & Briesch, A. M. (2017). Classification accuracy and acceptability of the Integrated Screening and Intervention System Teacher Rating Form. *School Psychology Quarterly, 32*(2), 212–225. https://doi.org/10.1037/spq0000147

De Ayala, J. R. (2022). *The theory and practice of item response theory.* Guilford Press.

*Deighton, J., Tymms, P., Vostanis, P., Belsky, J., Fonagy, P., Brown, A., Martin, A., Patalay, P., & Wolpert, M. (2013). The development of a school-based measure of child mental health. *Journal of Psychoeducational Assessment, 31*(3), 247–257. https://doi.org/f4x2xd

De Los Reyes, A., & Epkins, C. C. (2023). Introduction to the special issue. A dozen years of demonstrating that informant discrepancies are more than measurement error: Toward guidelines for integrating data from multi-informant assessments of youth mental health. *Journal of Clinical Child & Adolescent Psychology, 52*(1), 1-18. https://doi.org/10.1080/15374416.2022.2158843

*Dever, B. V., Gallagher, E. K., Hochbein, C. D., Loukas, A., & Dai, C. (2017). Examining subtypes of behavioral/emotional risk using cluster analysis. *Journal of Psychoeducational Assessment, 35*(6), 628–632. https://doi.org/gbsc82

Drummond, T. (1994). *The Student Risk Screening Scale (SRSS)*. Josephine County Mental

    Health Program.

*Dowdy, E., Nylund-Gibson, K., Felix, E. D., Morovati, D., Carnazzo, K. W., & Dever, B. V.

    (2014). Long-term stability of screening for behavioral and emotional risk. *Educational

    and Psychological Measurement, 74*(3), 453–472. https://doi.org/mjmc

*Downs, A., Strand, P. S., Heinrichs, N., & Cerna, S. (2012). Use of the teacher version of the

    Strengths and Difficulties Questionnaire with German and American preschoolers. *Early

    Education & Development, 23*(4), 493–516. https://doi.org/mjmd

Ebesutani, C., Reise, S. P., Chorpita, B. F., Ale, C., Regan, J., Young, J., Higa-McMillan, C., &

    Weisz, J. R. (2012). The revised child anxiety and depression scale-short version: Scale

    reduction via exploratory bifactor modeling of the broad anxiety factor. *Psychological

    Assessment, 24*(4), 833–845. https://doi.org/10.1037/a0027283

Eckert, T. L., Hintze, J. M., & Shapiro, E.S. (1999). Development and refinement of a measure

    for assessing the acceptability of assessment methods: The assessment rating profile-

    revised. *Canadian Journal of School Psychology, 15*, 21-42. https://doi.org/dtbdpz

*Eklund, K., Kilgus, S., von der Embse, N., Beardmore, M., & Tanner, N. (2017). Use of

    universal screening scores to predict distal academic and behavioral outcomes: A

    multilevel approach. *Psychological Assessment, 29*(5), 486–499. https://doi.org/f96wtc

*Elliott, S. N., Davies, M. D., Frey, J. R., Gresham, F., & Cooper, G. (2018). Development and

    initial validation of a social emotional learning assessment for universal screening.

    Journal of Applied Developmental Psychology, 55, 39–51.

    https://doi.org/10.1016/j.appdev.2017.06.002

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 8*(4), 341–

349. https://doi.org/10.1037/1040-3590.8.4.341

Epstein, M. H. (2004). *Behavioral and Emotional Rating Scale: A strengths-based approach to assessment (2nd ed.).* PRO-ED.

Epstein, M.H., & Sharma, J. (1998). *Behavioral and Emotional Rating Scale: A strength-based approach to assessment.* PRO-ED.

Epstein, M.H., & Synhorst, L. (2009). *Preschool Behavioral and Emotional Rating Scale*. PRO-ED.

*Erdogan, S., & Ozturk, M. (2011). Psychometric evaluation of the Turkish version of the pediatric symptom checklist-17 for detecting psychosocial problems in low-income children. *Journal of Clinical Nursing, 20*(17–18), 2591–2599. https://doi.org/cc2csk

*Español-Martín, G., Pagerols, M., Prat, R., Rivas, C., Sixto, L., Valero, S., Artigas, M. S., Ribasés, M., Ramos-Quiroga, J. A., Casas, M., & Bosch, R. (2021). Strengths and Difficulties Questionnaire: Psychometric properties and normative data for Spanish 5- to 17-year-olds. *Assessment, 28*(5), 1445–1458. https://doi.org/10.1177/1073191120918929

*Feil, E. G., & Becker, W. C. (1993). Investigation of a multiple-gated screening system for preschool behavior problems. *Behavioral Disorders, 19*(1), 44–53. https://doi.org/10.1177/019874299301900101

*Feil, E. G., Walker, H. M., & Severson, H. H. (1995). The Early Screening Project for young children with behavior problems. *Journal of Emotional and Behavioral Disorders, 3*(4), 194–202. https://doi.org/10.1177/106342669500300401

Flanagan, D. P., Alfonso, V. C., Primavera, L. H., Povall, L., & Higgins, D. (1996). Convergent validity of the BASC and SSRS: Implications for social skills assessment. *Psychology in the Schools*, *33*(1), 13–23. https://doi.org/dfqn9h

Forness, S., Freeman, S., Paperella, T., Kauffman, J., & Walker, H. (2012). Special education

implications of point and cumulative prevalence for children with emotional and

behavioral disorder. *Journal of Emotional and Behavioral Disorders*, *20*, 4–18.

https://doi.org/10.1177/1063426611401624

Gardner, W., & Kelleher, M. (1999). Pediatric Symptom Checklist-17 (PSC-17). *The Journal of

Pediatrics, 109*(2), 371–378. https://doi.org/10.1016/s0022-3476(86)80408-5

*Garrido, L. E., Barrada, J. R., Aguasvivas, J. A., Martínez-Molina, A., Arias, V. B., Golino, H.

F., Legaz, E., Ferrís, G., & Rojo-Moreno, L. (2020). Is small still beautiful for the

strengths and difficulties questionnaire? novel findings using exploratory structural

equation modeling. *Assessment, 27*(6), 1349–1367. https://doi.org/gdqbs4

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Multilevel reliability estimation within a

multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91.

https://doi.org/10.1037/e698612011-001

*Gillé, V., Kerkhoff, D., Heim-Dreger, U., Kohlmann, C. W., Lohaus, A., & Eschenbeck, H.

(2021). Stress-symptoms and well-being in children and adolescents: factor structure,

measurement invariance, and validity of English, French, German, Russian, Spanish, and

Ukrainian language versions of the SSKJ scales. *Health Psychology and Behavioral

Medicine, 9*(1), 875–894. https://doi.org/10.1080/21642850.2021.1990062

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening

assessments. *Journal of School Psychology, 45*(2), 117–135. https://doi.org/bj59gd

*Gomez, R., & Stavropoulos, V. (2020). Malaysian parent ratings of the Strengths and

Difficulties Questionnaire: Factor structure and measurement invariance across language

versions, gender, informants, and race. *Assessment, 27*(8), 1971–1984.

https://doi.org/10.1177/1073191118787284

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. J*ournal of Child Psychology and Psychiatry, 40*, 791–801.

Goodman, R., Iervolino, A. C., Collishaw, S., Pickles, A., & Maughan, B. (2007). Seemingly minor changes to a questionnaire can make a big difference to mean scores: a cautionary tale. *Social Psychiatry and Psychiatric Epidemiology*, *42*(4), 322–327. https://doi.org/10.1007/s00127-007-0169-0

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System (SSIS).* Pearson Assessment.

Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). Cross-informant agreement for ratings for social skill and problem behavior ratings: An investigation of the social skills improvement system—rating scales. *Psychological Assessment*, *22*(1), 157–166. https://doi.org/10.1037/a0018124

Gresham, F. M., Hunter, K. K., Corwin, E. P., & Fischer, A. J. (2013). Screening, assessment, treatment, and outcome evaluation of behavioral difficulties in an RTI model. *Exceptionality*, *21*(1), 19–33. https://doi.org/10.1080/09362835.2013.750115

*Gustafsson, B. M., Gustafsson, P. A., & Proczkowska-Björklund, M. (2016). The Strengths and Difficulties Questionnaire (SDQ) for preschool children—a Swedish validation. *Nordic Journal of Psychiatry, 70*(8), 567–574. https://doi.org/10.1080/08039488.2016.1184309

*Harrell-Williams, L. M., Raines, T. C., Kamphaus, R. W., & Dever, B. V. (2015). Psychometric analysis of the BASC–2 Behavioral and Emotional Screening System (BESS) student form: Results from high school student samples. *Psychological Assessment, 27*(2), 738–743. https://doi.org/10.1037/pas0000079

*Hartman, K., Gresham, F. M., & Byrd, S. (2017). Student internalizing and externalizing

behavior screeners: Evidence for reliability, validity, and usability in elementary schools.

Behavioral Disorders, 42(3), 108–118. https://doi.org/10.1177/0198742916688656

Health Resources and Services Administration. (2020). *National Survey of Children's Health*

Mental and Behavioral Health, 2018-2019. [Issue Brief].

https://mchb.hrsa.gov/sites/default/files/mchb/data-research/nsch-data-brief-mental-bh-

2019.pdf

Huang, F. L. (2018). Multilevel modeling myths. *School Psychology Quarterly*, *33*(3), 492–

499. https://doi.org/10.1037/spq0000272

*Huang, F. L., Reinke, W. M., Thompson, A., & Herman, K. C. (2019). An investigation of the

psychometric properties of the early identification system–student report. *Journal of*

Psychoeducational Assessment, 37(4), 473–485.

https://doi.org/10.1177/0734282918758791

Hysing, M., Elgen, I., Gillberg, C., Lie, S. A., & Lundervold, A. J. (2007). Chronic physical

illness and mental health in children. results from a large-scale population study. *Journal*

of Child Psychology and Psychiatry, *48*(8), 785–792. https://doi.org/10.1111/j.1469-

7610.2007.01755.x

*Iaccarino, S., von der Embse, N., & Kilgus, S. (2019). Interpretation and use of the social,

academic, and emotional behavior risk screener: A latent transition approach. *Journal of*

Psychoeducational Assessment, 37(4), 486–503. https://doi.org/gd32j9

*Isolan, L., Salum, G. A., Osowski, A. T., Amaro, E., & Manfro, G. G. (2011). Psychometric

properties of the screen for child anxiety related emotional disorders (scared) in Brazilian

children and adolescents. *Journal of Anxiety Disorders*, *25*(5), 741–748.

https://doi.org/10.1016/j.janxdis.2011.03.015

Jellinek, M. S., Murphy, J. M., & Burns, B. J. (1986). Brief psychosocial screening in outpatient

pediatric practice. *The Journal of Pediatrics, 109*(2), 371–378.

https://doi.org/10.1016/s0022-3476(86)80408-5

Jenkins, L. N., Demaray, M. K., Wren, N. S., Secord, S. M., Lyell, K. M., Magers, A. M.,

Setmeyer, A. J., Rodelo, C., Newcomb-McNeal, E., & Tennant, J. (2014). A critical

review of five commonly used social-emotional and behavioral screeners for elementary

or secondary schools. *Contemporary School Psychology*, *18*(4), 241–254.

https://doi.org/10.1007/s40688-014-0026-6

Kamphaus R. W., & Reynolds C. R. (2007). *BASC-2 Behavioral and Emotional Screening*

System Manual. Pearson.

Kamphaus, R. W., & Reynolds, C. R. (2015). *Behavior assessment system for children – third*

edition (BASC-3): Behavioral and emotional screening system (BESS). Pearson.

Kamps, D. M., Wills, H. P., Greenwood, C. R., Thorne, S., Lazo, J. F., Crockett, J. L.,

McGonigle Akers, J., & Swaggart, B. L. (2003). Curriculum influences on growth in

early reading fluency for students with academic and behavioral risks. *Journal of*

Emotional and Behavioral Disorders, *11*(4), 211–224. https://doi.org/ctntr6

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational*

Measurement, 50(1), 1–73. https://doi.org/10.1111/jedm.12000

Kennedy, C. H. (2005). *Single-case designs for educational research*. Pearson Education.

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2015).

A systematic review of evidence for the psychometric properties of the Strengths and

Difficulties Questionnaire. *International Journal of Behavioral Development*, *40*(1), 64–

75. https://doi.org/10.1177/0165025415570647

Keyes, C. (2006). Mental health in adolescence: Is America's youth flourishing? *American

Journal of Orthopsychiatry, 76*(3), 395–402. https://doi.org/10.1037/0002-9432.76.3.395

Kilgus, S. P., Chafouleas, S. M., & Riley-Tillman, T. C. (2013). Development and initial

validation of the Social and Academic Behavior Risk Screener for Elementary Grades.

School Psychology Quarterly, 28(3), 210–226. https://doi.org/10.1037/spq0000024

*Kilgus, S. P., Izumi, J. T., von der Embse, N. P., Van Wie, M. P., Eklund, K., Taylor, C. N., &

Iaccarino, S. (2019). Co-occurrence of academic and behavioral risk within elementary

schools: Implications for universal screening practices. *School Psychology, 34*(3), 261–

270. https://doi.org/10.1037/spq0000314

*Kilgus, S. P., Sims, W. A., von der Embse, N. P., & Taylor, C. N. (2016). Technical adequacy

of the Social, Academic, and Emotional Behavior Risk Screener in an elementary sample.

Assessment for Effective Intervention, 42(1), 46–59.

https://doi.org/10.1177/1534508415623269

*Kilgus, S. P., Sims, W. A., von der Embse, N. P., & Riley-Tillman, T. C. (2015). Confirmation

of models for interpretation and use of the Social and Academic Behavior Risk Screener

(SABRS). *School Psychology Quarterly, 30*(3), 335–352. https://doi.org/f7qj78

*Kilgus, S. P., von der Embse, N. P., Allen, A. N., Taylor, C. N., & Eklund, K. (2018).

Examining SAEBRS technical adequacy and the moderating influence of criterion type

on cut score performance. *Remedial and Special Education, 39*(6), 377–388.

https://doi.org/10.1177/0741932517748421

*Kilpatrick, K. D., Maras, M. A., Brann, K. L., & Kilgus, S. P. (2018). Universal screening for

social, emotional, and behavioral risk in students: DESSA-mini risk stability over time

and its implications for screening procedures. *School Psychology Review, 47*(3), 244–

257. https://doi.org/10.17105/spr-2017-0069.v47-3

*Kim, J., & Kamphaus, R. W. (2018). Investigation of factor structure and measurement

invariance by gender for the behavioral and emotional screening system among high school

students. *Psychological Assessment, 30*(2), 231–240. https://doi.org/10.1037/pas0000469

*King, K. R., Lembke, E. S., & Reinke, W. M. (2016). Using latent class analysis to identify

academic and behavioral risk status in elementary students. *School Psychology Quarterly,

31*(1), 43–57. https://doi.org/10.1037/spq0000111

King, K. R., & Reschly, A. L. (2014). A comparison of screening instruments: Predictive validity

of the BESS and BSC. *Journal of Psychoeducational Assessment*, *32*(8), 687–698.

https://doi.org/10.1177/0734282914531714

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis

of the effects of classroom management strategies and classroom management programs

on students' academic, behavioral, emotional, and motivational outcomes. *Review of

Educational Research, 86*(3), 643–680. https://doi.org/10.3102/0034654315626799

*Lambert, M. C., Epstein, M., Ingram, S., Simpson, A., & Bernstein, S. (2014). Psychometrics

and Measurement Invariance of the Emotional and Behavioral Screener. *Behavioral

Disorders, 39*(2), 89–101. https://doi.org/10.1177/019874291303900205

Lambert, M. C., January, S.-A. A., Epstein, M. H., Spooner, M., Gebreselassie, T., & Stephens,

R. L. (2015). Convergent validity of the behavioral and emotional rating scale for youth

in Community Mental Health Settings. *Journal of Child and Family Studies*, *24*(12),

3827–3832. https://doi.org/10.1007/s10826-015-0191-x

*Lambert, M. C., January, S. A. A., & Pierce, C. D. (2018). Latent structure of scores from the

> Emotional and Behavioral Screener. *Journal of Psychoeducational Assessment, 36*(3),

> 249–260. https://doi.org/10.1177/0734282916676130

Lambert, M. C., Sointu, E. T., & Epstein, M. H. (2018). A comprehensive review of international

> research using the behavioral and emotional rating scale. *International Journal of School

> & Educational Psychology*, *7*(3), 215–226. https://doi.org/gqtwhw

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical

> data. *Biometrics*, *33*, 159-174.

Lane, K. L. (2019). *Systematic Screening Tools: Universal Behavior Screeners*. Center on PBIS,

> University of Oregon. https://www.pbis.org/resource/systematic-screening-tools-
> universal-behavior-screeners

*Lane, K. L., Bruhn, A. L., Eisner, S. L., & Robertson Kalberg, J. (2010). Score Reliability and

> Validity of the Student Risk Screening Scale: A psychometrically sound, feasible tool for

> use in urban middle schools. *Journal of Emotional and Behavioral Disorders, 18*(4),

> 211–224. https://doi.org/10.1177/1063426609349733

Lane, K. L., Kalberg, J. R., & Menzies, H. M. (2009). *Developing schoolwide programs to

> prevent and manage problem behaviors: A step-by-step approach*. Guilford Press.

Lane, K. L., Kalberg, J. R., Menzies, H., Bruhn, A., Eisner, S., & Crnobori, M. (2011). Using

> systematic screening data to assess risk and identify students for targeted supports:

> Illustrations across the K-12 Continuum. *Remedial and Special Education*, *32*(1), 39–54.

> https://doi.org/10.1177/0741932510361263

Lane, K. L., & Menzies, H. M. (2009). *Student Risk Screening Scale for Internalizing and

> Externalizing (SRSS-IE)*. https://www.ci3t.org/screening

Lane, K. L., & Menzies, H. M. (2010). Student Risk Screening Scale for Early Childhood

     (SRSS-EC). https://www.ci3t.org/screening

Lane, K. L., Menzies, H. M, Oakes, W. P., & Kalberg, J. R. (2012). *Systematic screenings of*

     behavior to support instruction: From preschool to high school. Guilford Press.

Lane, K. L., Menzies, H. M., Oakes, W. P., & Kalberg, J. R. (2020). *Developing a schoolwide*

     framework to prevent and manage learning and behavior problems. Guilford Press.

Lane, K. L., & Oakes, W. P. (2010). *Screening Tool Rating Scale*. Available on ci3t.org.

Lane, K. L., Oakes, W. P., Buckman, M. M., Lane, N. A., Lane, K. S, Fleming, K., Swinburne

     Romine, R., Sherod, R., Chang, C., Jones, J., Cantwell, E., & Crittenden, M. (2023).

     Examination of the factor structure and measurement invariance of the SRSS-IE.

     Remedial and Special Education. https://doi.org/10.1177/07419325231193147

*Lane, K. L., Oakes, W. P., Ennis, R. P., & Royer, D. J. (2015). Additional evidence of

     convergent validity between SRSS-IE and SSiS-PSG scores. *Behavioral Disorders,*

     40(4), 213-229. https://doi.org/10.17988/0198-7429-40.4.213

Lane, K. L., Oakes, W. P., & Menzies, H. M. (2021). Considerations for systematic screening

     PK-12: Universal screening for internalizing and externalizing behaviors in the COVID-

     19 era. *Preventing School Failure: Alternative Education for Children and Youth, 65(*3),

     275-281. https://doi.org/10.1080/1045988X.2021.1908216

Lane, K. L., Oakes, W. P., Menzies, H. M., Buckman, M. M., & Royer, D. J. (2020). *Systematic*

     screening for behavior: Considerations and commitment to continued inquiry [research

     brief]. Ci3T Strategic Leadership Team. www.ci3t.org/screening

Lane, K. L., Oakes, W. P., Monahan, K. Smith, A., Lane, K. S., Buckman, M. M., Lane, N. A, &

     Sherod, R. (2023). A comparison of DESSA-mini and SRSS-IE screening tools.

Education and Treatment of Children. https://doi.org/10.1007/s43494-023-00106-9

Lane, K. L., Powers, L., Oakes, W.P., Buckman, M. M., Sherod, R. L., & Lane, K. S. (2020).

    Universal Screening – Systematic Screening to Shape Instruction: Lessons Learned &

    Practicalities. Center on PBIS, University of Oregon.

    https://www.pbis.org/resource/universal-screening-systematic-screening-to-shape-

    instruction

*Lane, K. L., Richards-Tutor, C., Oakes, W. P., & Connor, K. (2014). Initial evidence for the

    reliability and validity of the student risk screening scale with elementary age English

    learners. *Assessment for Effective Intervention, 39*(4), 219–232. https://doi.org/gcmc7h

Lebrun-Harris, L. A., Ghandour, R. M., Kogan, M. D., & Warren, M. D. (2022). Five-year trends

    in US children's health and well-being, 2016-2020. *JAMA Pediatrics*, 176(7), e220056-

    e220056. https://doi.org/10.1001/jamapediatrics.2022.0056

*Liau, A. K., Chow, D., Teck Kiang Tan, & Senf, K. (2011). Development and validation of the

    Personal Strengths Inventory using exploratory and confirmatory factor analyses. *Journal*

    of Psychoeducational Assessment, 29(1), 14–26.

    https://doi.org/10.1177/0734282910365648

*Liau, A. K., Tan, T. K., Li, D., & Khoo, A. (2012). Factorial invariance of the Personal

    Strengths Inventory-2 for children and adolescents across school level and gender.

    European Journal of Psychology of Education, 27(4), 451–465.

    https://doi.org/10.1007/s10212-011-0088-z

*Liu, J., Burgess, Y., DiStefano, C., Pan, F., & Jiang, N. (2020). Validating the Pediatric

    Symptoms Checklist–17 in the preschool environment. *Journal of Psychoeducational*

    Assessment, 38(4), 460–474. https://doi.org/10.1177/0734282919828234

Lohaus, A., Eschenbeck, H., Kohlmann, C. W., & Klein-Heßling, J. (2006). *Fragebogen zur Erhebung von Stress und Stressbewältigung im Kindes-und Jugendalter (SSKJ 3-8)*.

Ma, Z., Sherod, R. L., Lane, K. L., Buckman, M. M., & Oakes, W. P. (2022). *Interpreting universal behavior screening data: Questions to consider*. Center on PBIS, University of Oregon. https://www.pbis.org/resource/interpreting-universal-behavior-screening-data-questions-to-consider

*Margherio, S. M., Evans, S. W., & Owens, J. S. (2019). Universal screening in middle and high schools: Who falls through the cracks? *School Psychology, 34*(6), 591–602. https://doi.org/10.1037/spq0000337

Marzocchi, G. M., Capron, C., Di Pietro, M., Duran Tauleria, E., Duyme, M., Frigerio, A., Gaspar, M. F., Hamilton, H., Pithon, G., Simões, A., & Thérond, C. (2004). The use of the Strengths and Difficulties Questionnaire (SDQ) in southern European countries. *European Child & Adolescent Psychiatry*, *13*(S2), ii40–ii46. https://doi.org/10.1007/s00787-004-2007-1

McCoach, D. B., & Cintron, D. (2022). *Introduction to modern modelling methods*. Sage Quantitative Research.

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument Development in the Affective Domain: School and Corporate Applications*. Springer Science & Business Media.

McDougal, J., Bardos, A. N., & Meir, S. (2011). *Behavior Intervention Monitoring Assessment System*. Edumetrisis.

McDougal, J., Bardos, A. N., & Meir, S. (2016). *Behavior Intervention Monitoring Assessment System 2*. Edumetrisis.

McIntosh K., & Goodman, S. (2016). *Integrating multi-tiered systems of support: Blending RTI*

and PBIS. Guilford Press.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*(1), 114–140. https://doi.org/10.1037/met0000078

Merrell, K. W. (1993). Using behavior rating scales to assess social skills and antisocial behavior in school settings: Development of the school social behavior scales. *School Psychology Review, 22*(1), 115-133. https://doi.org/10.1080/02796015.1993.12085641

Merrell, K. W. (2002). *School Social Behavior Scales second edition*. Paul H. Brookes Publishing Co.

*Merrell, K. W., Felver-Gant, J. C., & Tom, K. M. (2011). Development and validation of a parent report measure for assessing social-emotional competencies of children and adolescents. *Journal of Child and Family Studies, 20*(4), 529–540. https://doi.org/10.1007/s10826-010-9425-0

*Mieloo, C. L., Bevaart, F., Donker, M. C. H., van Oort, F. V. A., Raat, H., & Jansen, W. (2014). Validation of the SDQ in a multi-ethnic population of young children. *European Jornal of Public Health, 24*(1), 26–32. https://doi.org/10.1093/eurpub/ckt100

Moulton, S. E., & Young, E. L. (2021). An item analysis of gender differences on the SRSS-IE in middle school. *Psychology in the Schools*, *58*(5), 789–803. https://doi.org/mjmp

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1),

143. https://doi.org/10.1186/s12874-018-0611-x

Murano, D., Sawyer, J. E., & Lipnevich, A. A. (2020). A meta-analytic review of preschool social and emotional learning interventions. *Review of Educational Research, 90*(2), 227– 263. https://doi.org/10.3102/0034654320914743

Muyskens, P., Marston, D. & Reschly, A.L. (2007) The use of response to intervention practices for behavior: An examination of the validity of a screening instrument. *The California School Psychologist 12*, 31–45 (2007). https://doi.org/10.1007/BF03340930

Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. B. (2011). *Devereux Student Strengths Assessment-mini (DESSA-mini).* Aperture Education.

National Center on Intensive Intervention. (2022). *Behavior screening tools chart rating rubric*. https://intensiveintervention.org/sites/default/files/call-tools/b-assessment/bs_rating_rubric_2022.pdf

*Naser, S., Brown, J., & Verlenden, J. (2018). The utility of Universal Screening to Guide School-Based Prevention Initiatives: Comparison of office siscipline referrals to standardized emotional and behavioral risk screening. *Contemporary School Psychology, 22*(4), 424–434. https://doi.org/10.1007/s40688-018-0173-2

*Naser, S., Hitti, A., & Overstreet, S. (2018). The behavioral and emotional screening system student form: Is there evidence of a global at-risk factor in a sample of predominantly African American youth? *Journal of Psychoeducational Assessment, 36*(5), 446–463. https://doi.org/10.1177/0734282916686015

Oakes, W. P., Buckman, M. M., Lane, K., L., & Sherod, R. L. (2021). *Selecting a universal behavior screening tool: Questions to consider.* Center on PBIS, University of Oregon.

https://www.pbis.org/resource/selecting-a-universal-behavior-screening-tool-questions-to-consider

Oakes, W. P., Lane, K. L., Cantwell, E. D., & Royer, D. J. (2017). Systematic screening for behavior in K–12 settings as regular school practice: Practical considerations and recommendations. *Journal of Applied School Psychology*, *33*(4), 369–393. https://doi.org/10.1080/15377903.2017.1345813

*Oakes, W. P., Lane, K. L., & Ennis, R. P. (2016). Systematic screening at the elementary level: Considerations for exploring and installing universal behavior screening. *Journal of Applied School Psychology, 32*(3), 214–233. https://doi.org/mjmq

Oakes, W. P., Lane, K. L., Ma, Z., Sherod, R. L., & Pérez-Clark, P. (2022). *Installing a universal behavior screening tool: Questions to consider*. Center on PBIS, University of Oregon. https://www.pbis.org/resource/installing-a-universal-behavior-screening-tool-questions-to-consider

Office of Special Education and Rehabilitative Services, U.S. Department of Education. (2022). *44th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act*. U.S. Department of Education. https://sites.ed.gov/idea/files/44th-arc-for-idea.pdf

*Pagano, M. E., Cassidy, L. J., Little, M., Murphy, J. M., & Jellinek, A. M. (2000). Identifying psychosocial dysfunction in school-age children: The pediatric symptom checklist as a self-report measure. *Psychology in the Schools*, 37(2), 91–106. https://doi.org/b7ksps

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E.,

McDonald, S., & McGuinness, L. A. (2021). The PRISMA 2020 statement: an updated

   guideline for reporting systematic reviews. *Systematic Reviews, 10*(1).

   https://doi.org/10.1186/s13643-021-01626-4

Pianta R. C. (1990) *Social Competence Scale*. University of Virginia.

*Pierce, C. D., Lambert, M., & Alamer, H. (2016). Convergent, criterion and social validity of

   the emotional and behavioral screener. *Journal of Child and Family Studies, 25*(1), 77–

   85. https://doi.org/10.1007/s10826-015-0206-7

Putnam, S. P., & Rothbart, M. K. (2006). Children's Behavior Questionnaire--Very Short Form

   (CBQ-Very Short). *APA PsycTests*. https://doi.org/10.1037/t07623-000

*Reinke, W. M., Herman, K. C., Huang, F., McCall, C., Holmes, S., Thompson, A., & Owens, S.

   (2022). Examining the validity of the Early Identification System–Student Version for

   screening in an elementary school sample. *Journal of School Psychology, 90*, 114-134.

   https://doi.org/10.1016/j.jsp.2021.10.001

Renshaw T. L. (2015). A replication of the technical adequacy of the Student Subjective

   Wellbeing Questionnaire. *Journal of Psychoeducational Assessment, 33,* 757-768.

*Renshaw, T. L. (2019). Screening Using the SDQ Total Difficulties Scale: An Analog Test of

   Three Cutoff Values. *Journal of Psychoeducational Assessment, 37*(8), 1030–1036.

   https://doi.org/10.1177/0734282918809808

*Renshaw, T. L., & Cook, C. R. (2018). Initial Development and Validation of the Youth

   Internalizing Problems Screener. *Journal of Psychoeducational Assessment, 36*(4), 366–

   378. https://doi.org/10.1177/0734282916679757

*Renshaw, T. L., & Cook, C. R. (2019). Preliminary psychometrics of responses to the Youth

   Externalizing Problems Screener. *Journal of Psychoeducational Assessment, 37*(8),

1016–1022. https://doi.org/10.1177/0734282918809814

Revelle, W. & Condon, D. M. (2019). Reliability from alpha to omega: A tutorial. *Psychological Assessment, 31* (12), 1395-1411. https://doi.org/10.1037/pas0000754

Reynolds C. R., & Richmond B. O. (2008). *Revised Children's Manifest Anxiety Scale–Second Edition (RCMAS-2)*. Western Psychological Services.

Sabey, C. V., Charlton, C. T., Pyle, D., Lignugaris-Kraft, B., & Ross, S. W. (2017). A review of classwide or universal social, emotional, behavioral programs for students in kindergarten. *Review of Educational Research, 87*(3), 512–543. https://doi.org/gfgtr3

*Schatschneider, C., Lane, K. L., Oakes, W. P., & Kalberg, J. R. (2014). The Student Risk Screening Scale: Exploring Dimensionality and Differential Item Functioning. *Educational Assessment, 19*(3), 185–203. https://doi.org/10.1080/10627197.2014.934608

Schmiege, S. J., Masyn, K. E., & Bryan, A. D. (2017). Confirmatory latent class analysis. *Organizational Research Methods*, *21*(4), 983–1001. https://doi.org/gcwbhp

Schonour, S, D., Lane, K. L., Oakes, W. P., Sherod, R. L., & Buckman, M. M. (2022). *The whys and hows of screening: Frequently asked questions for families*. Center on PBIS, University of Oregon. https://www.pbis.org/resource/the-whys-and-hows-of-screening-frequently-asked-questions-for-families

*Sharma, M., Chheda, S., Piramal, R., Bhatia, N., Frazier, T., & Singh, N. C. (2022). The social and emotional learning and orientation scale - development and validation of a brief measure in Hindi. *Journal of Psychoeducational Assessment, 40*(5), 571–591. https://doi.org/10.1177/07342829221075517

Shojaei, T., Wazana, A., Pitrou, I., & Kovess, V. (2008). The strengths and difficulties questionnaire: Validation study in French school-aged children and cross-cultural

comparisons. *Social Psychiatry and Psychiatric Epidemiology*, *44*(9), 740–747. https://doi.org/10.1007/s00127-008-0489-8

Sijtsma, K., & Pfadt, J. M. (2021). Part II: On the use, the misuse, and the very limited usefulness of Cronbach's alpha: Discussing lower bounds and correlated errors. *Psychometrika*, 86(4), 843–860. https://doi.org/10.1007/s11336-021-09789-8

Sinclair, E., Del'homme, M., & Gonzalez, M. (1993). Systematic screening for preschool behavioral disorders. *Behavioral Disorders, 18*(3), 177–188. https://doi.org/10.1177/019874299301800307

*Splett, J. W., Raborn, A., Lane, K. L., Binney, A. J., & Chafouleas, S. M. (2017). Factor analytic replication and model comparison of the BASC-2 Behavioral and Emotional Screening System. *Psychological Assessment, 29*(12), 1543–1549. https://doi.org/mjmt

*Splett, J. W., Smith-Millman, M., Raborn, A., Brann, K. L., Flaspohler, P. D., & Maras, M. A. (2018). Student, teacher, and classroom predictors of between-teacher variance of students' teacher-rated behavior. *School Psychology Quarterly, 33*(3), 460–468. https://doi.org/10.1037/spq0000241

Sugai, G., & Horner, R. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy, 24*(1-2), 23–50. https://doi.org/10.1300/J019v24n01_03

Treyvaud, K. (2014). Best practice when using the strengths and difficulties questionnaire with extremely preterm children: Are two informants better than one? *Developmental Medicine & Child Neurology*, *56*(5), 413–414. https://doi.org/10.1111/dmcn.12448

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., Weeks, L., Hempel, S. (2018). PRISMA extension for scoping

reviews (PRISMA-ScR): checklist and explanation. *Annals of Internal Medicine, 169*(7), 467-473. https://doi.org/10.7326/M18-0850

*Vancraeyveldt, C., Verschueren, K., Wouters, S., Van Craeyevelt, S., & Colpin, H. (2014). A multidimensional screening tool for preschoolers with externalizing behavior: Factor structure and factorial invariance. *Journal of Psychoeducational Assessment, 32*(8), 699–709. https://doi.org/10.1177/0734282914539898

*Vander Stoep, A., Mccauley, E., Thompson, K. A., Herting, J. R., Kuo, E. S., Stewart, D. G., Anderson, C. A., & Kushner, S. (2005). Universal emotional health screening at the middle school transition. *Journal of Emotional and Behavioral Disorders, 13*(4), 213–223. https://doi.org/10.1177/10634266050130040301

*Volpe, R. J., Yeung, T. S., Casale, G., Krull, J., Briesch, A. M., & Henneman, T. (2021). Evaluation of a German language school-based universal screening for student social, emotional, and behavioral risk. *International Journal of School & Educational Psychology, 9*(1), S10–S20. https://doi.org/10.1080/21683603.2020.1717699

*von der Embse, N. P., Kilgus, S. P., Iaccarino, S., & Levi-Nielsen, S. (2017). Screening for student mental health risk: Diagnostic accuracy, measurement invariance, and predictive validity of the Social, Academic, and Emotional Behavior Risk Screener-Student Rating Scale (SAEBRS-SRS). *School Mental Health, 9*(3), 273–283. https://doi.org/gbwd23

*von der Embse, N. P., Pendergast, L. L., Kilgus, S. P., & Eklund, K. R. (2016). Evaluating the applied use of a mental health screener: Structural validity of the Social, Academic, and Emotional Behavior Risk Screener. *Psychological Assessment, 28*(10), 1265–1275. https://doi.org/10.1037/pas0000253

Volpe, R. J., & Fabiano, G. A. (2013). *Daily behavior report cards: An evidence-based system of*

assessment and intervention. Guilford Press.

Volpe, R. J., Yeung, T. S., Casale, G., Krull, J., Briesch, A. M., & Henneman, T. (2021).

    Evaluation of a German language school-based universal screening for student social,

    emotional, and behavioral risk. *International Journal of School & Educational*

    Psychology, 9(1), S10–S20. https://doi.org/10.1080/21683603.2020.1717699

Walker, H. M., Forness, S. R., & Lane, K. L. (2014). Design and management of scientific

    research in applied school settings. In B. Cook, M. Tankersley, and T. Landrum (Eds.).

    Advances in Learning and Behavioral Disabilities (vol. 27, pp. 141-169). Emerald.

Walker, H. M., & Severson, H. H. (1992). *Systematic screening for behavior disorders* (SSBD).

    Sopris West.

Walker, H. M., Severson, H. H., & Feil, E. G. (1995). *Early screening project: A proven child-*

    find process. Sopris West.

Walrath, C. M., Mandell, D. S., Holden, E. W., & Santiago, R. L. (2004). Assessing the strengths

    of children referred for community-based mental health services. *Mental Health Services*

    Research, *6*(1), 1–8. https://doi.org/10.1023/b:mhsr.0000011252.84719.f2

*Warmbold-Brann, K., Maras, M. A., Splett, J. W., Smith-Millman, M., Dinnen, H., &

    Flaspohler, P. (2017). Examining the long-term stability of a strengths-based screener

    over 2 years. *Journal of Psychoeducational Assessment, 36*(8), 767–781.

    https://doi.org/10.1177/0734282917720564

Weist, M. D., Eber, L., Horner, R., Splett, J., Putnam, R., Barrett, S., Perales, K., Fairchild, A. J.,

    & Hoover, S. (2018). Improving multitiered systems of support for students with

    "internalizing" emotional/behavioral problems. *Journal of Positive Behavior*

    Interventions, *20*(3), 172–184. https://doi.org/10.1177/1098300717753832

Weist, M. D., Garbacz, A., Schultz, B., Bradshaw, C. P., & Lane, K. L. (2023). Revisiting the percentage of K-12 students in need of preventative interventions in schools in a "peri-COVID" era: Implications for implementing tiered programming. *Prevention Science* https://doi.org/10.1007/s11121-023-01618-x

Whalen, D. J., Sylvester, C. M., & Luby, J. L. (2017). Depression and Anxiety in Preschoolers: A Review of the Past 7 Years. *Child and Adolescent Psychiatric Clinics of North America*, *26*(3), 503–522. https://doi.org/10.1016/j.chc.2017.02.006

*Wiesner, M., & Schanding, G. T. (2013). Exploratory structural equation modeling, bifactor models, and standard confirmatory factor analysis models: Application to the BASC-2 Behavioral and Emotional Screening System Teacher Form. *Journal of School Psychology, 51*(6), 751–763. https://doi.org/10.1016/j.jsp.2013.09.001

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*(2), 203-14. https://doi.org/10.1901/jaba.1978.11-203

Wright, D. B., & Cook, C. R. (2014). *Brief Externalizing and Internalizing Screener for Youth*. Unpublished Instrument.

*Xu, W., Wang, M. C., Zhang, X., & Liu, Y. (2021). The Brief Problem Monitor (BPM-Y/BPM-P) Among Chinese Youth: Psychometric Properties and Measurement Invariance. *Journal of Psychopathology and Behavioral Assessment, 44*(2), 496–510. https://doi.org/10.1007/s10862-021-09927-7

Table 1

Identified Screening Instruments: Listed in Development Chronology

Name & Citation	Subscales	Rater(s)	Item Count	Response Option
Preschool Behavior Questionnaire (PBQ; Behar, 1977)	Hostile-Aggressive, Anxious-Fearful, Hyperactive-Distractible	Teacher, Parent	30	3-point Likert-type
Pediatric Symptom Checklist (PSC; Jellinek et al., 1986)	Internalizing, Externalizing, Problematic behavior, School problems	Parent, Youth	35	3-point Likert-type
Social Competence Scale - Parent Form (SCS-PF; Pianta, 1990)	Social Competence	Parent	16	4-point Likert-type
Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1992)	Stage 1: Externalizing & Internalizing Stage 2: Critical Events Index, Combined Frequency Index Stage 3: Observation (optional), School Archival Records Search (optional)	Teacher	S2: 56	Stage 1: ranking Stage 2: CEI - yes, no; CFI – 5-point Likert-type
Preschool Screening for Behavior Problems (PSBP; Feil & Becker, 1993)	Stage 1: Externalizing & Internalizing Stage 2: Critical Events Index, Combined Frequency Index Stage 3: Direct Observations	Teacher	S2: 48	Stage 1: ranking Stage 2: CEI - yes, no; CFI – 5-point Likert-type
School Social Behavior Scales (SSBS; Merrell, 1993)	Social Competence, Antisocial Behavior	Teacher	65	5-point Likert-type
Systematic Screening for Preschool Behavioral Disorders (Sinclair et al., 1993)	Stage 1: Externalizing & Internalizing Stage 2: Critical Events Index, Combined Frequency Index Stage 3: Observation (optional), School Archival Records Search (optional)	Teacher	S2: 56	Stage 1: ranking Stage 2: CEI - yes, no; CFI – 5-point Likert-type

Student Risk Screening Scale (SRSS; Drummond, 1994)	Antisocial Behavior	Teacher	7	4-point Likert-type
Early Screening Project (ESP; Walker et al., 1995)	Stage 1: Externalizing & Internalizing Stage 2: Critical Events Index, Combined Frequency Index Stage 3: Direct Observations	Teacher	S2: 42	Stage 1: ranking Stage 2: CEI - yes, no; CFI – 5-point Likert-type
Social Competence Scale (SCS; Conduct Problems Prevention Research Group, 1995)	Prosocial Behavior, Emotional Regulation Skills	Teacher	10	6-point Likert-type
Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher et al., 1997)	Panic/somatic, Generalized Anxiety, Separation Anxiety, Social Phobia, School Phobia	Student	41	3-point Likert-type
Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997)	Emotional Symptoms, Conduct Problems, Hyperactivity / Inattention, Peer Relationship Problems, Prosocial Behaviour	Teacher, Parent, Student	25	5-point Likert-type
Behavioral and Emotional Rating Scale (BERS; Epstein & Sharma, 1998)	Interpersonal Strength, Family Involvement, Intrapersonal Strength, School Functioning, Affective Strength	Teacher, Parent, Counselor, or Other persons knowledgeable about child	52	4-point Likert-type
Pediatric Symptom Checklist–17 (PSC-17; Gardner et al., 1999)	Internalizing Problems, Externalizing Problems, Attention Problems	Parents	17	3-point Likert-type
School Social Behavior Scales-2 (SSBS-2; Merrell, 2002)	Social Competence, Antisocial Behavior	Teacher	64	5-point Likert-type
Attention, Behavior, Language, and Emotions (ABLE) Universal Mental	Attention, Behavior, Language, Emotions	Teacher, Parent	22	yes/true, no/false, do not know;

Health Screening Tool (Barbarin, 2004)				
Behavioral and Emotional Rating Scale 2 (BERS-2; Epstein, 2004)	Interpersonal Strength, Family Involvement, Intrapersonal Strength, School Functioning, Affective Strength, Career Strength	Teacher, Parent, Student	Teacher: 52 Parent & Youth: 57	4-point Likert-type
Developmental Pathways Screening Questionnaire (DPSQ; Vander Stope et al., 2005)	Internalizing Emotional Distress, Externalizing Emotional Distress	Student	66 items	Not reported
Child Behavior Questionnaire (CBQ) - Very Short Form (Putnam & Rothbart, 2006)	Surgency, Negative Affectivity, Effortful Control	Teacher, Parent	36 items	7-point Likert-type
Mental Health Continuum (MHC) – Short Form (Keyes, 2006)	Emotional Well-being, Social Well-being, Psychological Well-being	Students	14 items	6-point Likert-type
Behavior Assessment System for Children–2nd Edition (BASC-2) Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007)	Personal Adjustment, Inattention/ Hyperactivity, Internalizing, School Problems	Teacher, Parent, Student	Teacher: 27 Parent: 30 Student: 30 Preschool: 25	4 point Likert-type
Behavior Screening Checklist (BSC; Muyskens et al., 2007)	Classroom Behaviors, Externalizing Behaviors, Socialization	Teacher	12	5-point Likert-type
Revised Children's Manifest Anxiety Scale, Second Edition (RCMAS-2) Short Form Total Anxiety Scale (Reynolds & Richmond, 2008)	Anxiety	Student	10	yes, no
Social Skills Improvement System – Performance Screening Guide (SSIS-PSG; Elliott & Gresham, 2008)	Prosocial Behavior, Motivation to Learn, Reading Skills, Math Skills	Teacher	4	5-level descriptive criterion

Preschool Behavioral and Emotional Rating Scale (PreBERS; Epstein & Synhorst, 2009)	Emotional regulation, School readiness, Social confidence, Family involvement	Teacher or Staff familiar with the child	42	4-point Likert-type
Student Risk Screening Scale-Internalizing and Externalizing (SRSS-IE; Drummond, 1994; Lane & Menzies, 2009)	Internalizing Behavior, Externalizing Behavior	Teacher	12	4-point Likert-type
Student Risk Screening Scale for Early Childhood (SRSS-EC; Lane & Menzies, 2010)	Internalizing Behavior, Externalizing Behavior	Teacher	11	4-point Likert-type
Achenbach System of Empirically Based Assessment - Brief Problem Monitor (ASEBA BPM; Achenbach et al. 2011)	Internalizing, Attention Problems, Externalizing, Total Problems	Teacher, Parent, Student	19	3-point Likert-type
Behavior Intervention Monitoring Assessment System (BIMAS; McDougal et al. 2011)	Conduct, Negative Affect, Cognitive/Attention, Social, Academic Functioning	Teacher, Parent, Student, Clinician	Teacher & Parent: 35 Student: 30 Clinician: 31	5-point Likert-type
Devereux Student Strengths Assessment Mini (DESSA-Mini; Naglieri et al., 2011)	Social-emotional Competence	Teacher, Parent, Counselor, or Other persons knowledgeable about the child	8	5-point Likert-type
Personal Strengths Inventory (PSI; Laui et al., 2011)	Emotional Awareness, Emotional Regulation, Goal Setting, Empathy, Social Competence	Student	22	5-point Likert-type
Social-Emotional Assets and Resiliency/Resilience Scales (SEARS; Merrell, 2011)	Self-regular/responsibility, Social Competence, Empathy	Teacher, Parent, Student	54	4-point Likert-type

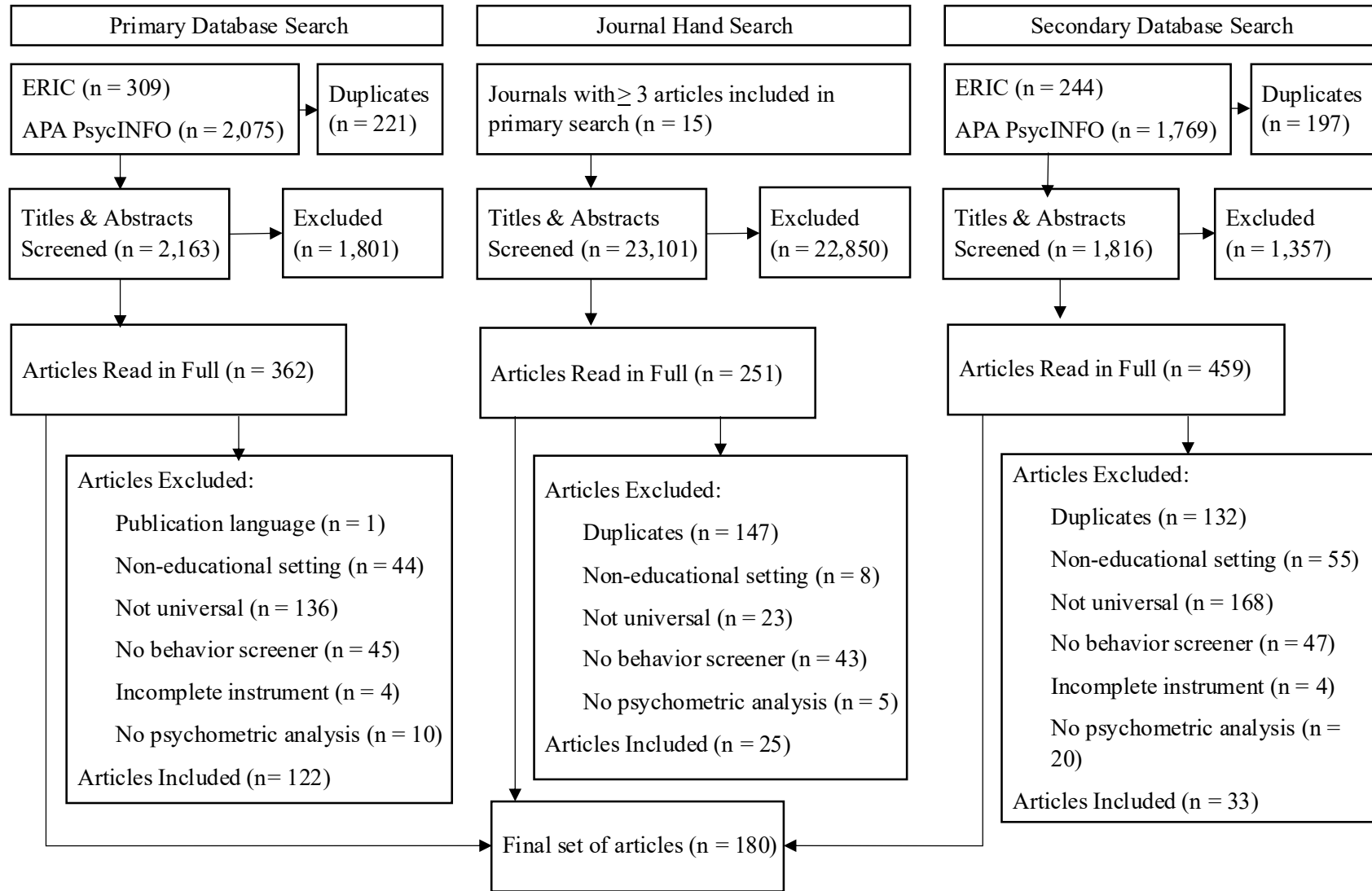|				
Student Internalizing Behavior Screener (SIBS; Cook et al., 2011)	Internalizing	Teacher	7	4-point Likert-type
Personal Strengths Inventory-2 (PSI-2; Laui et al., 2012)	Emotional Awareness, Emotional Regulation, Goal Setting, Empathy, Social Competence	Student	21	4-point Likert-type
Revised Child Anxiety and Depression Scale - Short Version (RCADS-25; Ebesutani et al., 2012)	Anxiety, Depression	Student	25	4-point Likert-type
Student Externalizing Behavior Screener (SEBS; Cook et al., 2012)	Externalizing	Teacher	7	4-point Likert-type
Direct Behavior Rating Single Item Scales (DBR-SIS; Chafouleas et al., 2013)	Disruptive, Academically Engaged, Respectful	Teacher	3	11-points
Emotional and Behavioral Screener (EBS; Cullian & Epstein, 2013)	Externalizing, Internalizing	Teacher, Other adult familiar with student	10	4-point Likert-type
Integrated Screening and Intervention System Teacher Rating Form (ITRF; Volpe & Fabiano, 2013)	Oppositional/Disruptive, Academic Productivity/Disorganization	Teacher	43	Stage 1: nomination Stage 2: ranking Stage 3: 4-point Likert-type
Me and My School Questionnaire (M&MS; Deighton et al., 2013)	Emotional Difficulties, Behavioral Difficulties	Student	16	3-point Likert-type
Social and Academic Behavior Risk Screener (SABRS; Kilgus et al., 2013)	Social Behavior, Academic Behavior	Teacher	12	4-point Likert-type
Brief Externalizing and Internalizing Screener for Youth (BEISY; Wright & Cook, 2014)	Externalizing, Internalizing	Teacher	6	5-point Likert-type

Hyperactivity, Opposition, Physical Aggression (HOPPA; Vancraeyveldt et al., 2014)	Hyperactivity, Opposition, Physical Aggression	Teachers	13	4-point Likert-type
Systematic Screening for Behavior Disorders – 2nd edition (SSBD; Walker et al., 2014)	Stage 1: Externalizing & Internalizing Stage 2: Critical Events Index, Combined Frequency Index	Teacher	S2: 56	Stage 1: ranking Stage 2: CEI - yes, no; CFI – 5-point Likert-type
Behavior Assessment System for Children–Third Edition (BASC-3) Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2015)	Personal Adjustment, Self-regulation, Internalizing	Teacher, Parent, Student	28	4-point Likert-type
Student Subjective Well-Being Questionnaire (SSWQ; Renshaw et al., 2015)	Joy of Learning, Educational Purpose, School Connectedness, Academic Efficacy	Student	16	4-point Likert-type
Behavior Intervention Monitoring Assessment System (BIMAS-2; McDougal et al., 2016)	Self- Awareness, Self-Management, Social Awareness, Relationship Skills, Responsible Decision Making	Teacher, Parent, Student	34	5-point Likert-type
Social, Academic, & Emotional Behavioral Risk Screener (SAEBRS; Kilgus, et al., 2016)	Social Behavior, Academic Behavior, Emotional Behavior	Teacher, Student	19	4-point Likert-type
Social Emotional Learning Assessment (SELA; Elliott, 2018)	Social Emotional Composite, Academic Functioning Composite	Teacher	8	5-level descriptive criterion
Stress and Coping Questionnaire for Children and Adolescents (SSKJ 3–8 R; Lohaus et al., 2018)	Somatic Symptoms, Anger, Sadness, Anxiety, Well-Being	Student	22	3-point Likert-type

Youth Internalizing Problems Screener (YIPS; Renshaw & Cook, 2018)	Internalizing	Student	10	4-point Likert-type
Early Identification System (EIS; Huang et al., 2019)	Externalizing Behavior, Internalizing Behavior, Peer Relations, School Engagement, Emotional Regulation, Relational Aggression	Teacher, Student	30	4-point Likert-type
Youth Externalizing Behavior Screener (YEBS; Arslan, 2019)	Conduct Problems, Hyperactivity, Attention Problems	Student	12	4-point Likert-type
Youth Externalizing Problems Screeners (YEPS; Renshaw & Cook, 2019)	Externalizing	Student	10	4-point Likert-type
Youth Internalizing Behavior Screener (YIBS; Arslan, 2021)	Anxiety, Depression	Student	10	4-point Likert-type
Social Emotional Learning and Orientation Scale (SELOS; Sharma et al., 2022)	Orientations, Challenges	Student	8	3-point Likert-type

Note. The exact number of items may vary across publications.

Figure 1

Overview of Search Process

Primary Database Search	Journal Hand Search	Secondary Database Search

ERIC (n = 309)
APA PsycINFO (n = 2,075) → Duplicates (n = 221)

Journals with ≥ 3 articles included in primary search (n = 15)

ERIC (n = 244)
APA PsycINFO (n = 1,769) → Duplicates (n = 197)

Titles & Abstracts Screened (n = 2,163) → Excluded (n = 1,801)

Titles & Abstracts Screened (n = 23,101) → Excluded (n = 22,850)

Titles & Abstracts Screened (n = 1,816) → Excluded (n = 1,357)

Articles Read in Full (n = 362)

Articles Read in Full (n = 251)

Articles Read in Full (n = 459)

Articles Excluded:
  Publication language (n = 1)
  Non-educational setting (n = 44)
  Not universal (n = 136)
  No behavior screener (n = 45)
  Incomplete instrument (n = 4)
  No psychometric analysis (n = 10)
Articles Included (n= 122)

Articles Excluded:
  Duplicates (n = 147)
  Non-educational setting (n = 8)
  Not universal (n = 23)
  No behavior screener (n = 43)
  No psychometric analysis (n = 5)
Articles Included (n = 25)

Articles Excluded:
  Duplicates (n = 132)
  Non-educational setting (n = 55)
  Not universal (n = 168)
  No behavior screener (n = 47)
  Incomplete instrument (n = 4)
  No psychometric analysis (n = 20)
Articles Included (n = 33)

Final set of articles (n = 180)

Figure 2

Cumulative Total of Universal Screening Articles by Year Published