Running Head: WORK AND SOCIAL ADJUSTMENT SCALE FOR YOUTH

The Work and Social Adjustment Scale for Youth: Psychometric Properties of the Teacher Version and Evidence of Contextual Variability in Psychosocial Impairments

Andres De Los Reyes[1]	Clayton R. Cook[2]	Margaret Sullivan[2]
Nicole Morrell[2]	Mo Wang[3]	Frank M. Gresham[4]
Corinne Hamlin[2]	Bridget A. Makol[1]	Lauren Keeley[1]
	Noor Qasmieh[1]	

[1]University of Maryland at College Park

[2]Univeristy of Minnesota

[3]University of Florida

[4]Louisiana State University

CITATION

De Los Reyes, A., Cook, C.R., Sullivan, M., Morrell, N., Hamlin, C., Wang, M., Gresham, F.M., Makol, B.M., Keeley, L.M., & Qasmieh, N. (2022). The Work and Social Adjustment Scale for Youth: Psychometric properties of the teacher version and evidence of contextual variability in psychosocial impairments. *Psychological Assessment*. Advance online publication. https://doi.org/10.1037/pas0001139

FUNDING INFORMATION

Efforts by the authors and the collection of data used for this study were supported by a grant from the Institute of Education Sciences (R324A180032; National Center for Special Education Research).

Author Notes

Author Contribution Statement: **Andres De Los Reyes, Ph.D.** (Conceptualization: Equal; Data curation: Equal; Formal analysis: Lead; Funding acquisition: Equal; Investigation: Equal; Methodology: Equal; Project administration: Equal; Supervision: Equal; Writing – original draft: Lead; Writing – review & editing: Lead). **Clayton R. Cook** (Conceptualization: Equal; Data curation: Equal; Formal analysis: Supporting; Funding acquisition: Equal; Methodology: Equal; Supervision: Equal; Writing – review & editing: Supporting). **Margaret Sullivan** (Data curation: Supporting; Project administration: Supporting; Writing – review & editing: Supporting). **Nicole Morrell** (Data curation: Supporting; Project administration: Supporting; Writing – review & editing: Supporting). **Mo Wang** (Conceptualization: Supporting; Formal analysis: Supporting; Funding acquisition: Supporting; Investigation: Supporting; Methodology: Supporting; Writing – review & editing: Supporting). **Frank M. Gresham** (Conceptualization: Equal; Funding acquisition: Equal; Methodology: Equal; Writing – review & editing: Supporting). **Corinne Hamlin** (Data curation: Supporting; Project administration: Supporting; Writing – review & editing: Supporting). **Bridget A. Makol** (Data curation: Supporting; Project administration: Supporting; Writing – review & editing: Supporting). **Lauren M. Keeley** (Data curation: Supporting; Project administration: Supporting; Writing – review & editing: Supporting). **Noor Qasmieh** (Data curation: Supporting; Project administration: Supporting; Writing – review & editing: Supporting)

Frank M. Gresham receives royalties on sales of the Social Skills Improvement System-Rating Scales, an instrument administered as part of this research.

Complete copies of study materials and analytic syntax are available upon request from the corresponding author, and a complete copy of the WSASY teacher version is available in online supplementary material. This study was not preregistered.

Correspondence regarding this manuscript should be addressed to Andres De Los Reyes, Comprehensive Assessment and Intervention Program, Department of Psychology, University of Maryland, Biology/Psychology Building, Room 3123H, College Park, MD 20742; Office: 301-405-7049; Fax: 301-314-9566; E-mail: adlr@umd.edu; Twitter: @JCCAP_Editor

Abstract

Effective mental health services require accurate assessment of psychosocial impairments linked to mental health concerns. Youth who experience these impairments do so within and across various contexts (e.g., school, home). Youth may display symptoms of mental health concerns without co-occurring impairments, and vice versa. Yet, nearly all impairment measures presume that those assessed display mental health concerns. Consequently, we recently developed youth and parent versions of a 5-item measure of youth psychosocial impairments (i.e., Work and Social Adjustment Scale for Youth [WSASY]), structured to assess any youth, regardless of mental health status. Across two studies, we developed and tested a WSASY teacher version, in a large sample of 382 student teacher reports (Study 1), and a subsample of 66 youth who, along with their parents and teachers, completed the WSASY and a series of school- and home-based behavioral tasks (Study 2). In Study 1, WSASY teacher reports demonstrated excellent internal consistency and unique relations with teacher reports on well-established measures of psychosocial strengths and difficulties. In Study 2, teacher, youth, and parent WSASY reports demonstrated low correspondence with each other and context-specific relations with criterion variables. This low correspondence allowed us to capitalize on an integrative approach designed to optimize informant-specific variance. Integrative scores demonstrated robust, large-magnitude relations with criterion variables across multiple information sources. These findings provide important psychometric support for use of WSASY teacher reports, and pave the way toward integrating WSASY reports from multiple informants who observe youth psychosocial impairments within different contexts and from different perspectives.

Keywords: Assessment; Impairment; Operations Triad Model; School; Work and Social Adjustment Scale

Public Significance Statement

Effectively addressing the needs of youth clients who receive mental health services requires accurately measuring how their behavior gets in the way or impairs their daily life. We evaluated a measure for assessing school-based impairments, and demonstrated a new strategy for accurately combining data from school-based and home-based impairment measures.

In mental health research and practice, effective delivery of services to children and adolescents (hereafter referred to collectively as "youth" unless otherwise specified) requires assessment of mental health symptoms as well as the degree to which these symptoms result in *psychosocial impairments* (American Psychiatric Association [APA], 2013). These impairments comprise examples of how mental health symptoms get in the way of, or interfere with, day-to-day functioning. As such, to assess these impairments accurately involves understanding *where* they manifest: the *contexts* of impairment. Among youth, impairments might manifest in a variety of contexts comprising their daily milieu, including the school setting, family environment, and interactions with peers (Fabiano & Pelham, 2016). Stated another way, mental health concerns among youth may result in impairments as diverse as decreases in school performance, family dysfunction, inability to meet demands at home (e.g., chores and homework), and/or difficulties with developing and maintaining relationships with peers. Thus, accurately assessing psychosocial impairments involves taking a holistic approach that incorporates data from a diverse set of information sources who have the collective *expertise* of observing these impairments across contexts. However, existing measures of psychosocial impairments are limited because they consist of only (a) single items, (b) items with leading questions that presume the presence of mental health concerns among those assessed, (c) items that are too broadly construed such that impairments could be accounted for by physical health concerns, and/or (d) a single source (e.g., clinician) whose ratings do not estimate impairments as they manifest within and across specific contexts. Two studies described in this paper fill a crucial gap in research on assessing youth psychosocial impairments, by testing the psychometric properties of a survey measure designed to assess school-based impairments (Study 1) and integrating these data with measures of home-based impairments (Study 2).

Because psychosocial impairments reflect the means by which mental health concerns interfere with life functioning, these impairments serve as "gateways" to receiving mental health services and monitoring how youth respond to these services. That is, the presence of impairments, in part, dictates crucial decisions germane to care, including making diagnoses, setting the goals of services, and monitoring response to services to determine whether they produce meaningful changes in youth functioning (Goldstein & Naglieri, 2016). As such, the path to developing precise, accurate measures of psychosocial impairments is a well-worn one, and a perennial priority in mental health research. Yet, an equally perennial concern presents itself with research on these measures: the relative paucity of well-conducted, rigorous research on the psychometric properties of impairment measures (Gadow et al., 2013; Rapee et al., 2012). The implications of this research gap cannot be understated. Impairment measures that lack strong psychometric properties result in inaccurate diagnoses, imprecise setting of service goals, and services that fail to achieve intended goals. Importantly, methodological issues surrounding widely used impairment measures lie in plain sight.

Four issues affect existing measures. First, the most widely used measures consist of single-item clinician ratings (e.g., Children's Global Assessment Scale, Global Assessment of Functioning; APA, 2013; Shaffer et al., 1983). Their brevity are overshadowed by their lack of psychometric rigor. Single-item measures often display variable and inconsistent precision and accuracy, such that their ability to consistently and accurately inform decision-making is rendered suspect (e.g., Blake et al., 2007; Tabachnick & Fidell, 2001).

Second, although multi-item measures exist, they typically take the form of diagnosis-specific measures of psychosocial impairments (e.g., Anxiety Disorders Interview Schedule for Children and Adolescents; Diagnostic Interview Schedule for Children; Costello et al., 1984;

Silverman & Albano, 1996). For instance, within diagnostic instruments, interviewers yoke assessments about impairments to specific diagnostic conditions. Although this allows for linking specific symptoms to specific impairments, youth who experience diagnosable mental health concerns often receive multiple diagnoses (see Goldstein & Naglieri, 2016). Relying on diagnosis-specific impairment measures creates interpretive issues, particularly given that (a) mental health concerns and psychosocial impairments are only moderately correlated (Allen et al., 2010; Storch et al., 2010) and (b) diagnostic thresholds fail to robustly differentiate youth on levels of psychosocial impairments (Egger et al., 2006; Keenan et al., 2010). Further, current research trends in service delivery involve developing services to address symptoms and impairments stemming from multiple, comorbid conditions (e.g., Weisz et al., 2012), and emerging models of classification seek to characterize mental health using concepts that traverse traditional diagnostic boundaries (e.g., Kotov et al., 2017). These issues create uncertainty with the long-term viability of diagnosis-specific modalities for assessing impairments.

Third, a key issue that cuts across many widely used measures—single-item and multi-item alike—lies in the very structure of the line of questioning in these measures. Each of the impairment measures noted above and other widely used measures (Impairment Rating Scale, Child Sheehan Disability Scale, Weiss Functional Impairment Rating Scale; Fabiano et al., 2006; Weiss et al., 2018; Whiteside, 2009) prompt respondents to make ratings based on the premise of existing mental health concerns (e.g., "How your child's problems affect his or her academic progress at school"). When items are framed this way, it precludes assessing youth until one establishes that they display symptoms. Such methods have obvious utility in service delivery settings. Yet, they have limited utility in screening contexts when the presence of symptoms is uncertain (e.g., unselected school-based samples; see also Bird et al., 2005).

Fourth, researchers have sought to address some of the limitations noted previously by creating *broad* impairment measures where areas of life interference are not tied to any kind of functioning (e.g., Columbia Impairment Scale; Brief Impairment Scale; Bird et al., 1996, 2005). However, by decoupling impairment ratings from displays of *psychosocial* functions, these measures allow informants to complete reports based on areas of impairment that result from a diffuse array of factors, including mental health concerns (e.g., anxiety, hyperactivity), but also physical conditions, and even life circumstances or adversities (e.g., parental divorce, neighborhood violence). Stated another way, framing items *too broadly* creates circumstances where variance in impairments could be attributed to factors beyond psychosocial functioning.

To address limitations of existing measures, we developed a short, multi-item measure that (a) could be completed by lay raters; (b) is freely available and amenable to wide dissemination; and (c) consists of items for which their framing disentangles mental health concerns from the presence of impairments but does so by retaining links to psychosocial functioning. These features facilitate not only assessing impairments linked to psychosocial functioning, but also in a way that is amenable to measuring any youth, regardless of mental health status.

The Work and Social Adjustment Scale for Youth (WSASY; De Los Reyes et al., 2019a) is a modified version of a widely used 5-item measure originally developed for use in assessing adults (WSAS; Mundt et al., 2002). The original WSAS assesses impairment across various domains (i.e., work, social life, home life, private life, close relationships), is sensitive to treatment response, and is capable of estimating naturalistic changes in impairment (e.g., Hussain et al., 2011; Kenwright et al., 2005; Kristensen et al., 2015; Zahra et al., 2014). The modified WSASY (a) assesses domains relevant to youth (e.g., completing chores and homework); (b) reflects impairments resulting from behavior generally, not mental health concerns in particular;

and (c) includes parallel versions for parents and youth to complete (see Appendices A and B of De Los Reyes et al., 2019a). That is, in contrast to instruments described previously for which item content is linked to symptoms (e.g., "How your child's problems affect…."), items on the WSASY are framed such that impairments are linked to psychosocial functioning but without reference to clinically elevated concerns (e.g., "Because of the ways I think, feel, or behave….").

In prior work (De Los Reyes et al., 2019a; Okuno et al., 2021), the WSASY displayed high internal consistency (i.e., α's $\geq .80$), particularly for a 5-item scale (see Youngstrom et al., 2019). In fact, scores from the WSASY parent and youth versions (a) relate to scores from a broad array of well-established survey measures of parent, youth, and family functioning; (b) distinguish adolescents on the number of co-occurring mental health concerns and peer-related impairments; and (c) predict independent observers' ratings of youths' interpersonal functioning in social interactions. Further, as with assessment of youth functioning generally (Achenbach et al., 1987; De Los Reyes et al., 2015, 2019b; Renk & Phares, 2004), parent and youth WSASY reports demonstrate low correspondence with one another ($r = .22$; De Los Reyes et al., 2019a).

Importantly, the existing WSASY versions do not include an informant with specific expertise in observing impairments in the school setting. Youth spend more of their time awake during weekdays in school than at home, and school performance is a critical social determinant of health that affects individuals over the course of their lifetime (Fabiano & Pelham, 2016). Thus, failure to assess psychosocial impairment in the school results in a significant information gap in accurately understanding *where* and *to what extent* youth display impairments.

Developing a WSASY teacher version opens doors for developing innovative strategies for integrating multi-informant WSASY reports. In particular, we require approaches that result in data that capture youth impairment across contexts. This is a crucial issue, as teacher reports of

youth functioning commonly disagree with the reports of parents and youth (for a review, see De Los Reyes et al., 2019c). Further, currently some researchers recommend addressing these *informant discrepancies* by examining informant's reports individually (e.g., Howe et al., 2019), resulting in increased Type I error. Others recommend using integrative or analytic strategies that assume informant discrepancies reflect measurement confounds (e.g., composite scores, tests of measurement invariance; see Olino et al., 2018; Sentse et al., 2009). These recommendations run counter to formal tests of these assumptions about informant discrepancies. Specifically, prior work finds that, when examined in relation to well-established, independent validation criteria (e.g., observed behavior, clinical indicators of treatment response), informant discrepancies often reflect *domain-relevant information*: Data that directly pertain to the domains about which informants provide reports (for reviews, see De Los Reyes et al., 2015, 2019b, 2022). In this respect, optimizing the accuracy of multi-informant assessments to predict relevant outcomes requires use of informants who systematically vary across multiple perspectives (e.g., self vs. other) and contexts of observation (e.g., home vs. school; see also Kraemer et al., 2003). That is, accurately integrating multi-informant assessment data requires *capitalizing* on the discrepancies they produce, rather than treating them as measurement confounds (see Makol et al., 2020).

Purpose and Hypotheses

To this end, in this paper we report two studies. In Study 1, we tested the psychometric properties of the WSASY teacher version (see **online supplementary material**) in a large, school-based sample. In Study 2, we examined a subsample of this larger sample of teachers who, along with youth and parents, each made WSASY reports. We tested the degree to which these reports reflect context-sensitive indices of youth psychosocial impairments, with a particular emphasis on domains of psychosocial functioning that predict academic engagement

and performance (see Cook et al., 2014; Reynolds et al., 2006). This multi-modal study included both multi-informant surveys of youth psychosocial difficulties (i.e., mental health concerns) and strengths (i.e., social skills), as well as independent observers' ratings of youth behavior in both classroom contexts and family interactions that typify the home context. Across studies, we tested (a) the internal consistency and criterion-related validity of scores taken from the WSASY teacher version (Study 1); (b) relations between teacher, youth, and parent WSASY versions and survey and behavioral indices of youth functioning (Study 2); and (c) an approach to integrating these WSASY reports (Study 2) that capitalizes on informants' unique perspectives and contexts to produce robust relations with criterion variables (Kraemer et al., 2003; Makol et al., 2020).

Study 1

Study 1 consisted of an initial test of the WSASY teacher version in a large, school-based sample. Consistent with prior work on the WSASY parent and youth versions, in this study we hypothesized that the WSASY teacher version would display relatively high internal consistency estimates (i.e., $\alpha \geq .80$; Nunnally & Bernstein, 1994), as well as unique relations to teachers' survey reports of youth psychosocial strengths and difficulties.

Method

Participants and Procedure

Teachers from three middle schools in a large, Midwestern metropolitan area were recruited to participate in this study through in-person meetings. Schools were racially (% non-white = 69%) and socioeconomically (% students receiving free and reduced lunch = 59%) diverse. We sent letters to parents informing them of the research study and providing an opportunity to opt out of participation. We recruited 35 teachers from the three schools, who completed surveys for youth who were randomly selected from a specified class period. With regard to demographics,

teachers were primarily female (66%) and white (91%). Teachers completed survey measures in one session and were provided with $10 per student packet completed in compensation for their time and effort. From a larger sample of 419 teacher reports, we included reports for 382 youth (School 1 = 107; School 2 = 127; School 3 = 148). Addressing our Study 1 aims hinged on obtaining internal consistency estimates of the WSASY teacher version. Thus, we included teachers in the final sample if they provided complete, item-level data on all constructs. The 382 teacher surveys in the final sample provided reports about psychosocial functioning of randomly selected youth enrolled in sixth (24.3%), seventh (33.2%), and eighth (33.5%) grade. These figures total below 100% because 8.9% of teachers did not provide grade data.

Measures

WSASY. We assessed youth psychosocial impairments using a modified version of the WSASY (De Los Reyes et al., 2019a), as described previously. The WSASY teacher version, which we include in online supplementary material, consisted of five items and reflected a modified version of the parent version with items revised to reflect teachers' perspectives (e.g., *"my child"* for parents vs. *"the student"* for teachers). Early in the study (i.e., for teachers in School 1), we administered the WSASY teacher version using the same item content as administered to parents. However, teachers in School 1 often encountered confusion with rating item 2, "Because of the ways this student thinks, feels, or behaves, her/his ability to complete household chores is impaired (for example, cleaning, tidying, helping with cooking, looking after brothers and sisters)." Thus, for Schools 2 and 3 we revised this item to reflect contexts about which teachers tend to display expertise in observing youth, "Because of the ways this student thinks, feels, or behaves, her/his ability to complete class assignments is impaired." The WSASY teacher version (see **online supplementary material**) reflects the version with the revised item

content. We observed no appreciable differences in Cronbach's alpha (α) estimates for WSASY teacher reports collected from School 1 versus Schools 2 and 3 and in fact, these estimates increased slightly for Schools 2 and 3 (i.e., αs = .84, .85, and .88, respectively). Further, removing item 2 from α calculations resulted in a *decrease* in α estimates at all three schools (i.e., αs = .79, .80, and .84, respectively). Each of these observations indicated that item 2 had no negative effects on internal consistency estimates and in fact contributed to favorable estimates, particularly for a 5-item scale (see also Youngstrom et al., 2019). Thus, for Study 1 we examined the total sample of teachers as one pooled sample. Teachers provided ratings on a scale of "0" (*not at all impaired*) to "8" (*very severely impaired*). Total scores could range from 0 to 40, with higher scores indicating greater impairment. We created measure instructions that prompted teachers to rate youth impairment resulting from youth behavior, without mention of mental health concerns. In this way, teachers in our sample could provide WSASY reports about youth impairment, regardless of the youth's mental health status.

Brief Problem Monitor (BPM). The BPM is a condensed version of the widely used Teacher Report Form from the Achenbach System of Empirically Based Assessments (ASEBA; Achenbach et al., 2011). The BPM consists of 18 items that collectively measure youth internalizing, attention, and externalizing concerns. Teachers provided ratings on a response scale with options of "0" (*not true*), "1" (*somewhat true*), and "2" (*very true*), wither greater scores indicating greater youth mental health concerns. Given the unselected nature of the sample, we aggregated teachers' ratings for each youth based on mean item responses for the 18 items. Teachers' BPM reports displayed relatively high internal consistency estimates, α = .92.

Social Skills Improvement System-Rating Scales (SSIS-RS). We used items from the SSIS-RS to assess youth social skills (Gresham & Elliott, 2008). The SSIS-RS is a widely used

measure of students' social skills and problem behaviors. The SSIS-RS includes common social skills across seven subdomains (i.e., strengths): Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, and Self-Control. The SSIS-RS has shown strong psychometric properties in terms of internal consistency, test-retest reliability, and validity estimates (Gresham et al., 2010). Teachers provided ratings on a response scale with options of "0" (*never*), "1" (*seldom*), "2" (*often*), and "3" (*almost always*), with greater scores indicating greater youth social skills. To minimize participant burden we used a 21-item subset of the SSIS-RS. As with the BPM, we aggregated teachers' ratings for each youth based on mean item responses for the 21 items. Teachers' reports on the 21 SSIS-RS items used in this study displayed relatively high internal consistency estimates, $\alpha = .97$. A complete list of the 21 SSIS-RS items used in this study is available upon request from the corresponding author.

Analytic Strategies

First, we calculated internal consistency (Cronbach's α) for all surveys. We also computed means and standard deviations for all surveys, and calculated statistics for skewness and kurtosis to determine if our data met assumptions for our parametric analyses (i.e., skewness/kurtosis in range of ±2.0; Tabachnick & Fidell, 2001). Second, we examined the nomological network of the WSASY (i.e., whether it is predicted by its theoretically relevant antecedents) by computing bivariate correlations among teachers' WSASY ratings and youth mental health concerns and social skills ratings. These correlations only provide information about whether the measures relate to each other at the bivariate level, and do not reveal whether scores taken from the survey measures of youth mental health concerns and social skills uniquely relate to WSASY teacher reports. Thus, we constructed a hierarchical linear regression to examine these unique relations. With WSASY teacher reports as the criterion variable, we entered scores from the BPM and

SSIS-RS simultaneously in a single step as independent variables.[1] For all tests, we interpreted statistical significance of findings based on a *p* value cutoff of < .05. We examined magnitudes of effect sizes based on Cohen's (1988) *d* (small: 0.30; medium: 0.50; large: 0.80) and *r* (small: .10; medium: .30; large: .50). Complete copies of Study 1 materials and analytic syntax are available upon request from the corresponding author, and a complete copy of the WSASY teacher version is available in online supplementary material. This study was not preregistered.

Results and Discussion

Preliminary Analyses and Internal Consistency Estimates for the WSASY Teacher Version

As mentioned previously, teachers' BPM (*M item score* = 0.36; *SD* = 0.39) and SSIS-RS (*M item score* = 2.31; *SD* = 0.65) reports displayed excellent internal consistency (i.e., αs ≥ .90). WSASY teacher reports (*M* = 6.66; *SD* = 7.86) displayed relatively high internal consistency (α = .87), particularly for a 5-item measure (see Youngstrom et al., 2019). These estimates closely approximated those observed in prior psychometric work on the parent and youth versions of the WSASY (i.e., αs of .84 and .85, respectively; De Los Reyes et al., 2019a). Further, all survey measures displayed acceptable levels of skewness and kurtosis (i.e., scores < ±2.0).

As mentioned previously, the WSASY teacher version underwent some modifications over the course of the study. Further, we administered only a subset of the original SSIS-RS items, in an effort to minimize participant burden. Taken together, these elements of our research process introduced some uncertainty regarding the precision of our Study 1 instruments. Thus, as an additional check on the reliabilities of Study 1 measures, we calculated omega coefficients (ω, a

[1] The BPM and SSIS-RS demonstrated a relatively large, negative bivariate correlation with each other, $r = -.85$, $p < .001$, suggesting the potential for collinearity effects hindering the interpretability of our regression-based approach to examining unique relations between these variables and the WSASY teacher version. Thus, we calculated regression diagnostics for this model, which yielded variance inflation factor (VIF) values of 3.6 for each of the two predictor variables. These values fall under both conventional (i.e., VIF ≥ 10) and more conservative (i.e., VIF ≥ 4) thresholds for detecting collinearity issues in regression models (see Hair et al., 2013).

model-based estimate of internal consistency; Revelle & Condon, 2019). Ratings on the WSASY teacher version had high reliability in the total sample (ω = .90, 95% confidence interval = [.88, .92]). Omega coefficients of the WSASY teacher version for the three schools were .86, .90, and .90, respectively. Because we used different statements for item 2 in School 1 from the other two schools, we calculated omega coefficients without item 2 as well. Omega coefficients without item 2 were .83, .83, and .85 for three schools, respectively. Because ratings on the WSASY teacher version had high reliabilities in all three schools with or without item 2, ratings of the five-item scale were used and the three schools were analyzed as a whole sample. Similarly, teachers' BPM and SSIS-RS reports both had high reliabilities (BPM: ω = .93, 95% confidence interval = [.92, .94]; SSIS-RS: ω = .97, 95% confidence interval = [.96, .97]). All of these data support the acceptable levels of precision of our Study 1 instruments.

Relations with Survey Measures of Psychosocial Strengths and Difficulties

The WSASY teacher version displayed large-magnitude bivariate relations with BPM and SSIS-RS mean item scores, *r*s = .85 and -.81; *p*s < .001. In a hierarchical regression analysis using the plan described previously, both the BPM and SSIS-RS uniquely related to WSASY teacher reports. The regression model explained a significant and large amount of variance in WSASY teacher reports, $R^2 = .75$; $F(2, 379) = 583.45$; $p < .001$. Further, the BPM and SSIS-RS each uniquely related to the WSASY teacher reports, and in the expected directions. That is, increased BPM scores related to increased WSASY teacher scores, whereas decreased SSIS-RS scores related to increased WSASY teacher scores, *β*s = .58 and -.32, respectively, *p*s < .001.

Taken together, Study 1 provides, for the first time, support for the psychometric properties of the WSASY teacher version. Prior work testing the properties of the parent and youth versions of the WSASY observed relatively high internal consistency estimates (αs > .80) and evidence

for convergent validity across a host of well-established measures of various youth, parent, and family functioning domains (De Los Reyes et al., 2019a). Similarly, we found that the WSASY teacher version uniquely related to measures of youth psychosocial difficulties (i.e., mental health concerns) and strengths (i.e., social skills). Yet, some important limitations and considerations regarding our findings warrant comment. Specifically, we were unable to collect demographic information for individual youth, and only had such information at the school level. This limitation was somewhat offset, in that we randomly assigned youth for teachers to rate. Thus, we had some confidence that the youth assessed reflected the diverse demographic characteristics observed at the school level. Further, we observed rather large correlations between the WSASY teacher version and measures of youth psychosocial difficulties and strengths. These correlations could be attributed to a number of factors, most notably shared informant variance (i.e., teachers completed all measures). More broadly, these large correlations raise questions as to the degree to which our findings generalized to use of more rigorous study designs. Collectively, these limitations necessitated a study that tested the properties of the WSASY teacher version using a multi-informant, multi-modal study design that allowed us to conduct in-depth assessments of each youth in the sample.

Beyond the limitations of Study 1, we previously noted that WSASY youth and parent versions display relatively low levels of correspondence with one another, consistent with cross-informant correspondence estimates for youth psychosocial functioning generally (e.g., Achenbach et al., 1987; De Los Reyes et al., 2015, 2019b; Renk & Phares, 2004). Thus, although we advanced the literature by developing and testing a WSASY teacher version, there remains uncertainty as to how to examine, interpret, and integrate teachers' reports relative to WSASY reports completed by youth and parents. This comprised our rationale for Study 2.

Study 2

In Study 2, we tested a multi-informant approach to collecting WSASY reports, in a subset of the large, school-based sample recruited for Study 1. We took a multi-modal approach that included multi-informant surveys of youth mental health concerns and social skills, as well as independent observers' ratings of youth behavior in classroom contexts and laboratory-based simulations of home-based contexts. We tested two hypotheses. First, we expected to observe significant relations between teacher, youth, and parent WSASY versions and indices of youth functioning. Second, we hypothesized that an approach to integrating teacher, youth, and parent WSASY reports that capitalizes on informants' unique perspectives and contexts would display robust relations with criterion variables (Kraemer et al., 2003; Makol et al., 2020).

Method

Participants and Procedure

We recruited participants for Study 2 based on the original sample of 419 youth and 35 teachers described in Study 1. Specifically, in conjunction with recruiting teachers we contacted parents of youth about whom teachers provided reports. Parents of youth who were randomly selected in Study 1 were emailed an invitation to participate in Study 2. The invitation included a survey link to provide consent and complete the parent survey. Among parents who consented to involvement in the study and completed the parent survey, we scheduled a time for them and their youth to participate in a parent-child interaction task mimicking common discussions in the home context. With regard to demographics, 36.4% of youth were male, and 63.6% were female. Youth ranged in age from 11 to 14 years, with a mean age of 12.6 years. A majority of youth were Caucasian (58%), 15.2% were African-American, 3% were American-Indian or Alaska Native, 1.5% were Asian American or Asian, 1.5% were Middle Eastern or North African, 1.5%

identified as “Other Race”, and 20% identified as 2 or more races. With regard to parents, the majority of parents were female (91%). Parents ranged in age from 24 to 55, with a mean age of 43.7 years. A majority of parents were Caucasian (66.7%), 15.2% identified as African American or Black, 7.6% identified as Asian American or Asian, 3% identified as American Indian/Alaska Native, 1.5% identified as “Other Race”, 1.5% identified as Middle Eastern/North African, and 4.5% identified as 2 or more races. Parents’ educational status varied as well, as 28.8% held a bachelor’s degree, 28.8% held a Master’s, 10.6% held an advanced degree (PhD., JD, MD, etc.), 13.6% completed a high school diploma or GED, 6% held an associate’s or vocational degree, 9.1% attended some college without completing a degree, and 3% did not have a diploma. Parents varied in their marital status, with 65.2% married, 16.7% never married, 10.6% divorced, 4.5% living together, 1.5% separated, and 1.5% widowed. A majority of parents (60.6%) made $901+ per week. In terms of parents’ relationship to their child, 97% were the youth’s biological parent, 1.5% were adoptive parents, and 1.5% were other relatives.

Addressing our Study 2 aims hinged on being able to obtain independent assessments of youth behavior using ecologically valid tasks reflecting school- and home-based functioning. Thus, we included participants in the final sample if they provided data for at least one of these context-based tasks. The 66 teacher-youth-parent triads in the final sample provided information about psychosocial functioning of youth enrolled in sixth (24.2%), seventh (28.8%), and eighth (37.9%) grade. As with Study 1, these figures total below 100% because 9.1% of teachers did not provide grade data. All WSASY teacher data were provided as part of Study 1 procedures described previously. Parents independently completed survey measures online and received $25 in compensation for their time and effort. The youth survey measures and parent-child interaction tasks were completed in one in-person session at the research site. Research staff

administered the youth surveys and facilitated the parent-child interaction tasks. Parents received $50 for their participation in the interaction task and youth received $60 in compensation for completing the youth surveys and participation in the interaction task. Classroom observations of each youth were completed during one in-person observation of a full class period.

Assessing Classroom Behavior

We collected school-based observations of youth behavior using well-established procedures of naturalistic observation, with trained observers conducting their assessments during one class period, averaging 50 minutes. Duration of observations varied slightly due to occasional school-wide schedule modifications and differences in class length across schools. Teachers informed students prior to observations taking place that guests may be present during some periods to observe the classroom. At the beginning of the class period, observers positioned themselves to have the clearest possible view of the target student while ensuring they did not disrupt class. When possible, two observers attended each classroom observation (51 out of 60 total observations) so that inter-rater reliability estimates could be obtained. Because the same observers completed both home-based and school-based observations, details on training procedures and inter-rater reliability are discussed subsequently.

Assessing Home Behavior

We collected home-based observations of youth behavior based on a well-established laboratory task designed to yield ecologically valid reflections of youth behavior when interacting with parental caregivers. Parents and youth first provided information to research personnel on a widely used survey measure of parent-youth conflict (i.e., Issues Checklist; Prinz et al., 1979), which assesses for the presence of conflict surrounding various day-to-day life topics (e.g., completing chores and homework). Based on these reports, research personnel

identified topics for use in a widely used 15-minute discussion task (Donenberg & Weisz, 1997). Within a laboratory setting housed in the university where the study took place, this task involved the parent-youth dyad discussing and attempting to resolve a topic assigned by research personnel. The task began with research personnel providing each parent-youth dyad a "practice topic" to acclimate them to the task aims and scope, namely to discuss planning a vacation (3 minutes). For the portion of the discussion focused on the conflict topic(s), a research staff member introduced the topic to the parent and youth, elicited a brief description from each dyad member of how this topic arises between them, and then explained that they would like the parent and youth to come to a resolution on the details of the topic for 6 minutes. The research staff member left the room during the task and returned after a 6 minute duration. Two discussions surrounding conflict topics took place during the interaction task, one identified by the parent and the other by the child. Each administration of the task was video-recorded so that independent observers could make ratings of home behavior. Behavioral indices derived from discussion tasks relate to measures of youth functioning across multiple modalities (e.g., physiology, performance-based tasks, surveys; Ehrlich et al., 2012; Thomas et al., 2017, 2019).

Independent Observers' Ratings of Classroom and Home Behavior

Independent observers were predominantly graduate research assistants, with the addition of select undergraduates and post-baccalaureates with training in psychology. We provided training for both the school-based and home-based observations and behavioral coding over multiple sessions totaling 11 hours, not including independent practice and homework assignments. Group training procedures included didactic teaching of behavioral codes with discussion of specific scenarios to establish clarity, demonstrations of the parent-child interaction task, role-playing, and practice coding sample videos of both school-based and home-based scenarios.

Based on observations of behavior displayed by youth in the school- and home-based tasks, we trained independent observers to make behavioral ratings for use as criterion measures in tests of the validity of scores taken from informants' WSASY reports. Specifically, we applied a common coding scheme to assess youth behavior across the two tasks, with a response scale consisting of the options of "0" (*absent*), "1" (*somewhat present*), and "2" (*very present*). Trained observers made ratings using the "0-1-2" response scale described previously, separately by school and home behavior. Observers rated school behavior naturalistically (i.e., *in vivo* observations). Observers rated home behavior using videotapes of task administrations. We created a scheme of specific behavioral codes that observers could apply in parallel fashion, across observations of school and home behavior. Using this scheme, observers rated participants on 18 different behaviors—nine strengths and nine difficulties—within the broader categories of: (1) externalizing problems, (2) internalizing problems, (3) attention problems, (4) self-management skills, (5) relationship skills, and (6) social awareness. For some of these behaviors, we included a "no opportunity" option for observation periods in which the situation failed to present opportunities for the youth to display the rated behavior. For example, a rater would select "no opportunity" for the behavior "able to resolve conflicts/disagreements with others" if no conflict or disagreements occurred during the observation period.

Originally, we attempted to code youth behavior across a range of behaviors indicative of psychosocial strengths (e.g., calmness, listening to and helping others) as well as difficulties (e.g., disobedience, apathy, difficulty concentrating). However, we observed relatively low inter-rater reliability estimates for codes of psychosocial strengths, and we observed these low reliability estimates for both school behavior and home behavior ratings. We were only able to attain reliable codes for a set of observed youth difficulties. In particular, our assessment of

youth behavior focused on a set of observed difficulties across a range of youth concerns, namely behaviors indicative of disobedience, disruptive behavior, apathy/lack of motivation, and difficulty concentrating. To assess inter-rater reliability for the selected behaviors, we calculated intraclass correlation coefficients (*ICC*s) for both home-based and school-based observations. Specifically, we computed separate one-way random effects, absolute agreement, single rater models for home-based and school-based observations. We chose this model because different sets of raters rated different subjects (every rater did not rate all subjects), and because our basis of measurement uses single rater scores rather than the mean of scores (Koo & Li, 2016). These models produced average *ICC* estimates of .62 and .71 for home- and school-based codes, respectively. These estimates qualify as "good" according to commonly cited inter-rater reliability standards (Cicchetti, 1994). We created separate scores of school-based (M = 1.83; SD = 2.32) and home-based (M = 0.56; SD = 1.31) difficulties, and internal consistency estimates for both school-based and home-based scores were in the range deemed acceptable, αs = .84 and .70, respectively. Although the school-based ratings displayed acceptable levels of skewness and kurtosis (i.e., range of ±2.0), the home-based ratings did not. This likely reflected the observation that the home-based codes were zero-inflated, with 48 youth receiving a score of "0." To address this issue, we dichotomized home-based codes based on whether or not the youth received a score above zero. We used this dichotomized score in tests reported below.

Survey Measures

Teachers, youth, and parents, completed surveys of youth psychosocial functioning. Teachers and parents also completed forms to collect the demographic information described previously. Across survey measures, we made minor modifications to item content, to fit each informant's perspective (e.g., *"I"* for youth; *"my child"* for parents; *"the student"* for teachers).

WSASY. We administered the WSASY to teachers as described in Study 1, and to parents and youth as described previously (see also Appendices A and B of De Los Reyes et al, 2019a).

BPM. Teachers completed the BPM as described in Study 1. Parents and youth completed parallel versions of the BPM. The parent and youth versions of the BPM contain one more item than the teacher version (i.e., 19 items). As in Study 1, we computed mean item scores for all informants' reports, with higher scores indicating greater youth mental health concerns.

SSIS-RS. For Study 2, teachers' reports were based on their Study 1 data. Parents and youth completed parallel versions of the SSIS-RS. As in Study 1, we computed mean item scores for all informants' reports, with higher scores indicating greater youth social skills.

Analytic Strategies

Computing Descriptive and Reliability Statistics. We carried out computations of descriptive and internal consistency reliability statistics for all survey measures as described in Study 1. For our behavioral codes, we estimated inter-rater reliability via *ICC*s, using the procedures we previously described.

Integrating Multi-Informant WSASY Reports. As mentioned previously, we leveraged a multi-informant approach to integrating WSASY reports that included reports completed by teachers, youth, and parents. In multi-informant assessments, reports from these informants commonly yield discrepant estimates of youth functioning, including in school-based samples (for reviews, see De Los Reyes et al., 2015, 2019c). To optimize prediction of criterion variables using these discrepant data, we leveraged an integrative approach developed by Kraemer and colleagues (2003). Specifically, this approach involves repurposing traditional techniques for analyzing variations among survey items—principal components analysis (PCA; Nunnally & Bernstein, 1994)—to instead model multi-informant data. Following Kraemer and colleagues'

PCA approach, we synthesized the variability among informants' reports into sources informed by prior work (see Achenbach et al., 1987; De Los Reyes et al., 2013, 2015).

First, *context* variation reflects the environment in which the informant observes the youth. Second, *perspective* variation reflects whether the report comes from an observer or self-rater. Third, *trait* variation reflects concerns that manifest across informants' contexts and perspectives. We used a set of informants who collectively varied in their contexts and perspectives, with (a) informants observing from a school-based, observer perspective (teachers); (b) informants observing from a home-based, observer perspective (parents); and (c) informants observing from a self-perspective and based on a mix of home and school contexts (youth). Importantly, these were the same set of informants' reports used by Kraemer and colleagues, and we synthesized these reports using the same PCA-based modeling procedures. We determined the presence of these three sources of variability via examination of component weights, consistent with Kraemer and colleagues (2003). We expected our PCA to include components for *Trait* (i.e., all informants' reports load strongly and in the same direction), *Context* (i.e., informants from different contexts load in opposite directions), and *Perspective* (i.e., self-reports load in the opposite direction of observer informants' reports).

Consistent with prior work (Kraemer et al., 2003; Makol et al., 2020), we conducted an unrotated PCA on the three informants' WSASY reports. In this respect, our subject-to-item ratio (i.e., 61/3 = 20.33:1) meets the typical subject-to-item ratio deemed "large" within PCA modeling contexts (e.g., 20:1; see Osborne & Costello, 2004). Within this unrotated PCA, we set the number of components to be extracted to three. We examined principal component weights for each informant's report to determine whether we identified trait, context, and perspective scores described previously. Makol and colleagues found that the *Trait* score yielded optimal

prediction of criterion variables, an observation consistent with Kraemer and colleagues' notions as to the value of this approach. Thus, we used the WSASY *Trait* score for analyses reported below. We also computed bivariate correlations among all informants' WSASY reports.

Links between Multi-Informant WSASY Reports and Criterion Variables. To examine relations among WSASY reports and criterion variables we ran two separate sets of analyses. First, we computed bivariate correlations among the teacher, youth, and parent WSASY reports and the WSASY *Trait* score and all criterion variables that we could analyze with continuous data. For comparison purposes, we also calculated a *composite score* (i.e., the simple average of correlations between teacher, youth, and parent WSASY reports and criterion variables). In this way, we examined the "value-added" by the WSASY *Trait* score, in that we benchmarked the magnitude of *Trait* score correlations against a method of aggregating correlation data that, unlike the *Trait* score approach, assumes that variations among the different correlations (i.e., informant-specific variance) reflect measurement confounds (see also Makol et al., 2020).

Second, as mentioned previously, we observed a large proportion of youth in the sample (n = 48) who received observer ratings of "0" on home-based behavioral displays of difficulties. For these data, we computed a series of independent samples *t*-tests, with the independent variable being whether the youth displayed one or more difficulties during the task (0 = no difficulties; 1 = one or more difficulties), and the dependent variables being the teacher, youth, and parent WSASY reports and the WSASY *Trait* score. We also computed Cohen's *d* effect sizes for all tests, as well as simple averages of the effects based on teacher, youth, and parent WSASY reports (i.e., to compare this average effect to effects observed with the WSASY *Trait* score). Complete copies of Study 2 materials and analytic syntax are available upon request from the corresponding author. This study was not preregistered.

Results and Discussion

Preliminary Analyses

We report in Table 1 descriptive statistics for all of our survey measures. We conducted preliminary analyses to test if our data met assumptions for our planned parametric analyses (i.e., skewness/kurtosis in range of ±2.0). We examined frequency distributions for all variables used in analyses reported below to assess normality. Scores for most measures fell within acceptable ranges of skewness and kurtosis. However, there were a few exceptions. Specifically, WSASY youth and teacher reports, as well as BPM teacher and parent reports, exhibited relatively high skewness and/or kurtosis, which we would expect with a non-clinical sample of youth. Thus, we applied a square-root transformation to these scores. These transformed scores displayed skewness and kurtosis statistics within acceptable levels. All analyses reported below used these transformed scores. As seen in Table 1, across all informants' survey reports, we observed acceptable internal consistency estimates (i.e., αs $\geq$ .70; Nunnally & Bernstein, 1994), and as mentioned previously, *ICC* estimates for the home- and school-based ratings (i.e., .62 and .71, respectively) indicated "good" inter-rater reliability for these ratings (Cicchetti, 1994).

Integrating Multi-Informant WSASY Reports

In Table S1, we report statistics relevant to the PCA we used to integrate youth, teacher, and parent WSASY reports. Consistent with our expectations, an examination of the principal component weights for each informant's report supported our identification of trait, context, and perspective scores described previously. First, our PCA included a *Trait* score component in which all informants' reports loaded strongly and in the same direction (i.e., all factor loadings > .40). Second, we observed a *Context* score component, in which reports from teachers loaded strongly in one direction, reports from parents loaded strongly in the direction opposite of

teachers, and reports from youth did not load strongly in either direction (i.e., consistent with youth reports including a "mix" of home- and school-based observations). Third, we observed a *Perspective* score component in which youth self-reports loaded strongly and in the direction opposite of the two observer informants' reports (i.e., parent and teacher).

Beyond the factor loadings in Table 2, bivariate correlations among informants' WSASY reports revealed a low-to-moderate youth-teacher correlation ($r = .29$; $p < .05$), a non-significant youth-parent correlation ($r = .19$; $p = .13$), and a near-zero teacher-parent correlation ($r = .03$; $p = .82$). In contrast, the WSASY *Trait* score displayed moderate-to-large correlations with each of the WSASY reports from youth, teachers, and parents (rs = .81, .69, and .48, respectively; ps < .001). Thus, consistent with prior work using this integrative approach (Kraemer et al., 2003; Makol et al., 2020), these findings signify that the WSASY *Trait* score integrated multi-informant data in a way that preserved the unique information contributed by the informants, which all tended to display relatively low correlations among their reports.

Links with Survey Measures of Psychosocial Strengths and Difficulties

In Table 2, we report correlations between informants' WSASY reports and criterion variables reflecting survey measures of psychosocial strengths and difficulties, as well as correlations between the WSASY *Trait* score and these same criterion variables. Across informants, the general pattern reflected larger-magnitude within-informant correlations (i.e., WSASY reports and criterion variable reports completed by the same informant), relative to between-informant correlations (i.e., WSASY reports and criterion variable reports completed by different informants). When viewed through the traditional lens or interpretation of patterns of multi-trait, multi-method correlations, this pattern might signal a threat to the validity of scores taken from informants' WSASY reports (see Campbell & Fiske, 1959; Nunnally & Bernstein,

1994). That is, this traditional view hinges on two interpretations being true: (a) multiple findings from disparate methods of collecting data—in this case, multiple informants' reports—should point to the same conclusion, and thus (b) discrepancies among findings reflect measurement confounds (see also Olino et al., 2018). The problem with this interpretation is that does not withstand empirical scrutiny. Indeed, it lies in stark contrast with the last decade of research on informant discrepancies in assessments of youth functioning (see De Los Reyes et al., 2019b).

Specifically, the most well-conducted studies about informant discrepancies probe their meaning using *validation testing*, namely examining informant discrepancies in relation to independent assessments of behavior (for reviews, see De Los Reyes et al., 2019b, 2019c). These studies reveal that patterns of informant discrepancies reflect the notions that (a) youth vary in the contexts in which they display behavioral signs of psychosocial functioning and (b) informants systematically vary in where they observe youth (e.g., De Los Reyes et al., 2009; Deros et al., 2018). When data support these notions, approaches that seek to capitalize on systematic, valid variations among informants' reports should consequently yield more robust relations with criterion variables, relative to approaches that either leverage individual informants' reports, or aggregate data under the assumption that informant discrepancies hold no probative value (i.e., they reflect measurement confounds). These are the exact patterns of findings observed in Table 2. Specifically, the WSASY *Trait* score—an integrative approach that treats discrepancies among informants' reports as containing valid information about youth functioning (Kraemer et al., 2003; Makol et al., 2020)—yielded more robust effects *across* validity indicators, relative to any one individual informant's WSASY report or the simple average of effects across their reports. In subsequent analyses, we sought to find out whether these findings generalized to independent assessments of youth behavior.

Links with Observed Behavior

In Tables 2 and 3, we report findings relevant to links between informants' WSASY reports and criterion variables reflecting observed behavior, as well as links between the WSASY *Trait* score and these same criterion variables. The findings across Tables 2 and 3 provide important corroboration of findings described previously. Indeed, here too findings regarding links between informants' WSASY reports and observed behavior were largely informant-specific. Do these informant-specific effects reflect measurement confounds? Findings based on the WSASY *Trait* score rule out this possibility. As with the correlations with survey reports in Table 2, the WSASY *Trait* score robustly and at large-magnitude effects was related to school-based and home-based observations of youth difficulties.

General Discussion

Main Findings and Implications

Across two studies, we made a series of important findings that support recently developed assessments of youth psychosocial impairments. Specifically, we observed psychometric support for the WSASY—a short, multi-item survey measure of youth psychosocial impairments—when completed by teachers. As with the parent and youth versions (De Los Reyes et al., 2019a), this version of the WSASY related to various aspects of youth functioning collected across several measurement modalities including independent observations of behavior and other informants' survey reports. Yet, we also observed that WSASY teacher reports displayed relatively low levels of correspondence with WSASY reports completed by youth and parents. Some assume that these low levels of correspondence signal issues with the veracity of informants' reports on survey measures like the WSASY (e.g., Olino et al., 2018). Of course, this is merely an *assumption* about discrepancies among informants' reports of youth functioning. We subjected

this assumption to validation testing. By leveraging recent developments in strategies for integrating multi-informant assessments (Makol et al., 2020), we observed findings that contradict this assumption and at the same time, are consistent with the latest evidence on the meaning of informant discrepancies commonly observed in assessments of youth functioning (Achenbach et al., 1987; De Los Reyes et al., 2015, 2019b; Renk & Phares, 2004).

Not only do multiple informants' WSASY reports yield discrepant estimates of youth impairments, these discrepancies also signal a *strength* of this multi-informant approach. Consistent with Kraemer and colleagues' (2003) notions about multi-informant assessments of youth mental health, discrepancies among informants' reports of youth impairments appear to signal contextual variability in these impairments. In line with this notion, our findings support the need for future research to take a multi-informant approach to assessing youth impairments *and* use integrative strategies that capitalize on the discrepant estimates they reveal. Stated another way, various integrative approaches have as a key underlying assumption the idea that these discrepancies signal measurement confounds (e.g., composite scores, structural equations modeling, measurement invariance; De Los Reyes et al., 2019b, 2019c). Our findings indicate that these assumptions should be carefully examined, using multi-modal paradigms of validation testing. Indeed, if validation tests reveal that a multi-informant assessment contains discrepancies that reflect domain-relevant information—like those observed in Study 2—then the implications are clear: Use of approaches to integrating multi-informant data that violate these assumptions result in drawing invalid inferences from study findings. Taken together, our findings point to both the need to take a multi-informant approach to assessing youth impairments, as well as carefully select the approaches one takes to integrating multi-informant data.

Our findings also have important implications for measurement development and testing of

teacher-rated instruments. Indeed, recent, large-scale studies speak to the potential for implicit bias effects on the part of teacher reports about student functioning (e.g., Chin et al., 2020). These implicit bias effects impact our understanding of student functioning domains like psychosocial impairments in at least two important ways. First, implicit biases may impact teachers' perceived student impairments, such that they may depress the ability of teacher-rated instruments to accurately estimate such impairments, particularly those of students whose backgrounds differ from teachers' own backgrounds (Millsap, 2011). Second, implicit biases may directly impact student impairments, insofar as these biases may impact how teachers interact with students whose backgrounds differ from teachers' own backgrounds, thus potentially impacting student functioning (e.g., DeCuir-Gunby & Bindra, 2022). In these respects, implicit biases have the potential to create disparities in the accuracy of teacher-rated estimates of impairments, as well as contribute to disparities in psychosocial impairments. Our study did not include an index of implicit biases, and thus we were unable to test the degree to which implicit bias effects moderated relations between the WSASY teacher version and validity indicators. Needless to say, these issues merit further study.

Limitations and Future Directions

Some limitations of Study 1 and 2 warrant comment. First, we constructed the WSASY teacher version so that it would be developmentally appropriate for informants to complete about youth. However, our study is the first to examine this version's psychometric properties and did so using a sample of middle school youth. Thus, future research should test whether the psychometric properties of scores taken from this version of the WSASY generalize to samples of relatively younger and older children. Second, we tested the WSASY teacher version in a non-clinical sample of youth. We were able to demonstrate that teachers' WSASY reports related to

multiple measures of clinically relevant constructs. At the same time, we recommend that future studies test the psychometric properties of the WSASY in clinic settings (e.g., psychiatric hospitals and community mental health centers). Third, our tests of criterion-related validity focused on domains of psychosocial functioning that predict academic engagement and performance (see Cook et al., 2014; Reynolds et al., 2006). Yet, we were unable to include in our studies direct measures of academic performance. Further research ought to test the criterion-related validity of scores taken from the WSASY in relation to measures of such performance.

Fourth, our Study 2 sample was not only much smaller than our Study 1 sample, but it was also less diverse in terms of youth racial background. Granted, our Study 2 sample, though less diverse, nonetheless consisted of roughly 42% youth whose parents reported as having a non-Caucasian racial background. Further, the original parent and youth versions of the WSASY were developed and tested using a majority-minority sample of youth, with only a third of the sample identifying as White (De Los Reyes et al., 2019). In these respects, the WSASY teacher version as tested in this study produced comparable reliability estimates and, when compared to the parent and youth WSASY versions, produced superior estimates of score validity, particularly relative to parent reports (see Table 2). An additional consideration is that the PCA used to integrate multi-informant data in Study 2 produced factor loadings comparable to a more recent study that applied this sample PCA modeling approach to multi-informant data collected within a sample roughly double the size of Study 2 (Makol et al., 2020). This last observation is not surprising, given that patterns and levels of cross-informant correspondence manifest at remarkably robust, consistent levels across the globe, as indicated by a recent meta-analysis of studies conducted in over 30 countries that traverse all inhabited continents (De Los Reyes et al., 2019b). Nevertheless, we encourage further psychometric testing of the WSASY teacher

versions in more diverse and larger samples of youth.

Fifth, for our Study 2 behavioral coding procedures we could only arrive at a reliable set of codes for assessing youth difficulties. Importantly, prior work indicates that the WSASY parent and youth versions each relate to observed youth strengths (i.e., social skills; De Los Reyes et al., 2019a). Further, Study 2 indicated that the WSASY teacher version robustly related to all survey reports of youth strengths (i.e., youth, teacher, and parent; Table 2). Thus, a key area for future research involves examining whether links between scores taken from the WSASY and observed youth strengths extend to the WSASY teacher version.

Sixth, although we included multiple informants' reports about youth impairments and examined these reports across multiple youth functioning domains and assessment modalities, we did not include a secondary impairment measure. In this psychometric study of the WSASY, we wanted to test the measure in a sample of participants who varied as to the presence of mental health concerns. In this way, we were able to test a key question about the measure: Could we collect psychometrically sound data about psychosocial impairments, even among participants with low rates of clinically significant mental health concerns? By construction, our study could not include a secondary measure of impairment, because all current impairment measures require the presence of clinically elevated mental health concerns in order to collect impairment data. Thus, we encourage researchers to test the psychometric properties of the WSASY in clinical samples in which participants complete the WSASY and alternative impairment measures.

Concluding Comments

In both research and clinical work, sound decision-making requires accurately assessing psychosocial impairments linked to mental health concerns. Youth who experience impairments do so in various ways and across myriad contexts, and may display signs of these impairments

without co-occurring mental health concerns, and vice versa. Yet, nearly all impairment measures presume the presence of mental health concerns among those assessed. Further, we lack paradigms for integrating multi-informant data about youth impairments. In this study, we tested the WSASY teacher version, a 5-item measure of youth impairments that facilitates assessing any youth, regardless of mental health status. Across two studies, we learned that WSASY teacher reports displayed excellent internal consistency and demonstrated unique relations with teacher reports on well-established measures of psychosocial strengths and difficulties. Further, an integrative approach designed to capitalize on the unique, informant-specific data provided by teachers, youth, and parents demonstrated the most robust, large-magnitude relations across all criterion variables, regardless of the information source. These findings pave the way toward integrating WSASY reports from multiple informants who observe youth psychosocial impairments within different contexts and from different perspectives.

References

Achenbach, T. M., McConaughy, S. H., Ivanova, M. Y., & Rescorla, L. A. (2011). *Manual for the ASEBA Brief Problem Monitor (BPM)*. ASEBA.

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*, 213-232. https://doi.org/10.1037/0033-2909.101.2.213

Allen, J. L., Lavallee, K. L., Herren, C., Ruhe, K., & Schneider, S. (2010). DSM-IV criteria for childhood separation anxiety disorder: Informant, age, and sex differences. *Journal of Anxiety Disorders*, *24*, 946-952. https://doi.org/10.1016/j.janxdis.2010.06.022

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.) (*DSM-V*). American Psychiatric Association.

Bird, H. R., Andrews, H., Schwab-Stone, M., Goodman, S., Dulcan, M., Richters, J., Rubio-Stipec, M., Moore, R. E., Chiang, P.-H., Hoven, C., Canino, G., Fisher, P., & Gould, M. S. (1996). Global measures of impairment for epidemiologic and clinical use with children and adolescents. *International Journal of Methods in Psychiatric Research, 6*, 295-307. https://doi.org/10.1002/(SICI)1234-988X(199612)6:4<295::AID-MPR173>3.3.CO;2-5

Bird, H. R., Canino, G. J., Davies, M., Ramírez, R., Chávez, L., Duarte, C., & Shen, S. (2005). The Brief Impairment Scale (BIS): A multidimensional scale of functional impairment for children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*, 699-707. https://doi.org/10.1097/01.chi.0000163281.41383.94

Blake, K., Cangelosi, S., Johnson-Brooks, S., & Belcher, H. M. (2007). Reliability of the GAF and CGAS with children exposed to trauma. *Child Abuse and Neglect*, *31*, 909-915. https://doi.org/10.1016/j.chiabu.2006.12.017

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105. https://doi.org/10.1037/h0046016

Chin, M. J., Quinn, D. M., Dhaliwal, T. K., & Lovison, V. S. (2020). Bias in the air: A nationwide exploration of teachers' implicit racial attitudes, aggregate bias, and student outcomes. *Educational Researcher*, *49*, 566-578. https://doi.org/10.3102/0013189X20937240

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284-290. https://doi.org/10. 1037/1040-3590.6.4.284

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cook, C. R., Volpe, R. J., & Delport, J. (2014). Systematic progress monitoring of students with emotional and behavioral disorders: The promise of change-sensitive brief behavior rating scales. In H. M. Walker, & F. M. Gresham (Eds.). *Handbook of evidence-based practices for emotional and behavioral disorders: Applications in schools* (pp. 211-228). Guilford.

Costello, A. J., Edelbrock, C. S., Dulcan, M. K., & Kalas, R. (1984). Testing of the NIMH Diagnostic Interview Schedule for Children (DISC) in a clinical population (Contract No. DB-81-0027). Final report to the Center for Epidemiological Studies, National Institute of Mental Health. University of Pittsburg.

DeCuir-Gunby, J. T., & Bindra, V. G. (2022). How does teacher bias influence students?: An introduction to the special issue on teachers' implicit attitudes, instructional practices, and student outcomes. *Learning and Instruction*, *78*, 101523. https://doi.org/10.1016/j.learninstruc.2021.101523

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D., &

Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*, 858-900. https://doi.org/10.1037/a0038498

De Los Reyes, A., Makol, B. A., Racz, S. J., Youngstrom, E. A., Lerner, M. D., & Keeley, L. M. (2019a). The Work and Social Adjustment Scale for Youth: A measure for assessing youth psychosocial impairment regardless of mental health status. *Journal of Child and Family Studies*, *28*, 1-16. https://doi.org/10.1007/s10826-018-1238-6

De Los Reyes, A., Lerner, M. D., Keeley, L. M., Weber, R., Drabick, D. A. G., Rabinowitz, J., & Goodman, K. L. (2019b). Improving interpretability of subjective assessments about psychological phenomena: A review and cross-cultural meta-analysis. *Review of General Psychology*, *23*, 293-319. https://doi.org/10.1177/108926801983764

De Los Reyes, A., Cook, C. R., Gresham, F. M., Makol, B. A., & Wang, M. (2019c). Informant discrepancies in assessments of psychosocial functioning in school-based services and research: Review and directions for future research. *Journal of School Psychology*, *74*, 74-89. https://doi.org/10.1016/j.jsp.2019.05.005

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, *37*, 637-652. https://doi.org/10.1007/s10802-009-9307-3

De Los Reyes, A., Talbott, E., Power, T., Michel, J., Cook, C. R., Racz, S. J., & Fitzpatrick, O. (2022). The Needs-to-Goals Gap: How informant discrepancies in youth mental health assessments impact service delivery. *Clinical Psychology Review*, *92*, 102114. https://doi.org/10.1016/j.cpr.2021.102114

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical*

Psychology, 9, 123-149. https://doi.org/10.1146/annurev-clinpsy-050212-185617

Deros, D. E., Racz, S. J., Lipton, M. F., Augenstein, T. M., Karp, J. N., Keeley, L. M., . . . De Los Reyes, A. (2018). Multi-informant assessments of adolescent social anxiety: Adding clarity by leveraging reports from unfamiliar peer confederates**.** *Behavior Therapy*, *49*, 84-98. https://doi.org/10.1016/j.beth.2017.05.001

Donenberg, G. R., & Weisz, J. R. (1997). Experimental task and speaker effects on parent-child interactions of aggressive and depressed/anxious children. *Journal of Abnormal Child Psychology*, *25*, 367-387. https://doi.org/10.1023/A:1025733023979

Egger, H. L., Kondo, D., & Angold, A. (2006). The epidemiology and diagnostic issues in preschool attention-deficit/hyperactivity disorder: A review. *Infants and Young Children*, *19*, 109-122. https://doi.org/10.1097/00001163-200604000-00004

Ehrlich, K. B., Dykas, M. J., & Cassidy, J. (2012). Tipping points in adolescent adjustment: Predicting social functioning from adolescents' conflict with parents and friends. *Journal of Family Psychology*, *26*, 776-783. https://doi.org/10.1037/a0029868

Fabiano, G. A., Pelham, W. E., Waschbusch, D. A., Gnagy, E. M., Lahey, B. B., Chronis, A. M. et al. (2006). A practical measure of impairment: Psychometric properties of the Impairment Rating Scale in samples of children with attention-deficit/hyperactivity disorder and two school-based samples. *Journal of Clinical Child and Adolescent Psychology*, *35*, 369-385. https://doi.org/10.1207/s15374424jccp3503_3

Fabiano, G. A., & Pelham, W. E. (2016). Impairment in children. In S. Goldstein & J. A. Naglieri (Eds.), *Assessing Impairment* (pp. 105-119). NY: Springer.

Gadow, K. D., Kaat, A. J., & Lecavalier, L. (2013). Relation of symptom-induced impairment with other illness parameters in clinic-referred youth. *Journal of Child Psychology and*

Psychiatry, *54*(11), 1198-1207. https://doi.org/10.1111/jcpp.12077

Goldstein, S., & Naglieri, J. A. (2016). *Assessing impairment: From theory to practice*. Springer.

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System-Rating Scales.* Pearson.

Gresham, F. M., Elliott, S. N., & Kettler, R. J. (2010). Base rates of social skills acquisition/performance deficits, strengths, and problem behaviors: An analysis of the Social Skills Improvement System-Rating Scales. *Psychological Assessment, 22*(4)*,* 809-815. https://doi.org/10.1037/a0020255

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2013). *Multivariate data analysis: Advanced diagnostics for multiple regression* [Online supplement]. Retrieved from http://www.mvstats.com/Downloads/Supplements/Advanced_Regression_Diagnostics.pdf

Howe, G. W., Dagne, G. A., Brown, C. H., Brincks, A. M., Beardslee, W., Perrino, T., & Pantin, H. (2019). Evaluating construct equivalence of youth depression measures across multiple measures and multiple studies. *Psychological Assessment, 31*, 1154-1167. https://doi.org/10.1037/pas0000737

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, *3*, 29-51. https://doi.org/10.1146/annurev.clinpsy.3.022806.091419

Hussain, A., Weisaeth, L., & Heir, T. (2011). Psychiatric disorders and functional impairment among disaster victims after exposure to a natural disaster: a population based study. *Journal of Affective Disorders*, *128*, 135-141. https://doi.org/10.1016/j.jad.2010.06.018

Keenan, K., Wroblewski, K., Hipwell, A., Loeber, R., & Stouthamer-Loeber, M. (2010). Age of onset, symptom threshold, and expansion of the nosology of conduct disorder for girls. *Journal of Abnormal Psychology*, *119*, 689-698. https://doi.org/10.1037/a0019346

Kenwright, M., Marks, I., Graham, C., Franses, A., & Mataix-Cols, D. (2005). Brief scheduled

phone support from a clinician to enhance computer-aided self-help for obsecessive-compulsive disorder: Randomized control trial. *Journal of Clinical Psychology*, *61*, 1499-1508. https://doi.org/10.1002/jclp.20204

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M. et al. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*, 454-477. https://doi.org/10.1037/abn0000258

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry*, *160*(9), 1566-1577. https://doi.org/10.1176/appi.ajp.160.9.1566

Kristensen, P., Weisæth, L., Hussain, A., & Heir, T. (2015). Prevalence of psychiatric disorders and functional impairment after loss of a family member: A longitudinal study after the 2004 Tsunami. *Depression and Anxiety*, *32*, 49-56. https://doi.org/10.1002/da.22269

Makol, B. A., Youngstrom, E. A., Racz, S. J., Qasmieh, N., Glenn, L. E., & De Los Reyes, A. (2020). Integrating multiple informants' reports: How conceptual and measurement models may address long-standing problems in clinical decision- making. *Clinical Psychological Science*, *8*, 953-970. https://doi.org/10.1177/2167702620924439

Millsap, E. (2011). *Statistical methods for studying measurement invariance*. Taylor & Francis.

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social

Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*, *180*, 461-464. https://doi.org/10.1192/bjp.180.5.461

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill

Okuno, H., Rezeppa, T., Raskin, T., & De Los Reyes, A. (2021). Adolescent safety behaviors and social anxiety: Links to psychosocial impairments and functioning with unfamiliar peer confederates. *Behavior Modification*. Advance online publication. https://doi.org/10.1177/01454455211054019

Olino, T. M., Finsaas, M., Dougherty, L. R., & Klein, D. N. (2018) Is parent–child disagreement on child anxiety explained by differences in measurement properties? An examination of measurement invariance across informants and time. *Frontiers in Psychology, 9*, 1295. https://doi.org/10.3389/fpsyg.2018.01295

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research, and Evaluation*, *9*, Article 11. https://doi.org/10.7275/ktzq-jq66

Prinz, R. J., Foster, S. L., Kent, R. N., & O'Leary, K. D. (1979). Multivariate assessment of conflict in distressed and non-distressed mother–adolescent dyads. *Journal of Applied Behavior Analysis*, *12*, 691-700. https://doi.org/10.1901/jaba.1979.12-691

Rapee, R. M., Bögels, S. M., van der Sluis, C. M., Craske, M. G., & Ollendick, T. H. (2012). Annual research review: Conceptualising functional impairment in children and adolescence. *Journal of Child Psychology and Psychiatry*, *53*, 454-468. https://doi.org/10.1111/j.1469-7610.2011.02479.x

Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review*, *24*, 239-254.

https://doi.org/10.1016/j.cpr.2004.01.004

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

Reynolds, C. R., Livingston, R. B., & Willson, V. (2006). *Measurement and assessment in education.* Upper Saddle River, NJ: Pearson.

Sentse, M., Veenstra, R., Lindenberg, S., Verhulst, F. C., & Ormel, J. (2009). Buffers and risks in temperament and family for early adolescent psychopathology: Generic, conditional, or domain-specific effects? The TRAILS study. *Developmental Psychology*, *45*, 419-430. https://doi.org/10.1037/a0014072

Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry*, *40*, 1228-1231. https://doi.org/10.1001/archpsyc.1983.01790100074010

Silverman, W. K., & Albano, A. M. (1996). *The Anxiety Disorders Interview Schedule for DSM-IV-Child and Parent Versions*. Oxford University Press.

Storch, E. A., Larson, M. J., Muroff, J., Caporino, N., Geller, D., Reid, J. M. et al. (2010). Predictors of functional impairment in pediatric obsessive-compulsive disorder. *Journal of Anxiety Disorders*, *24*, 275-283. https://doi.org/10.1016/j.janxdis.2009.12.004

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Allyn and Bacon.

Thomas, S. A., Jain, A., Wilson, T., Deros, D., Jacobs, I., Ehrlich, K. B., Dunn, E., Aldao, A., & De Los Reyes, A. (2019). Moderated mediation of the link between parent-adolescent conflict and adolescent risk-taking: The role of physiological regulation and hostile behavior in an experimentally controlled investigation. *Journal of Psychopathology and Behavioral Assessment*, *41*, 699-715. https://doi.org/10.1007/s10862-019-09747-w

Thomas, S. A., Wilson, T., Jain, A., Deros, D. E., Um, M., Hurwitz, J., Jacobs, I., Myerberg, L., Ehrlich, K. B., Dunn, E. J., Aldao, A., Stadnik, R., & De Los Reyes, A. (2017). Toward developing laboratory-based parent-adolescent conflict discussion tasks that consistently elicit conflict-related stress responses: Support from physiology and observed behavior. *Journal of Child and Family Studies, 26*, 3288-3302. https://doi.org/10.1007/s10826-017-0844-z

Weiss, M. D., McBride, N. M., Craig, S., & Jensen, P. (2018). Conceptual review of measuring functional impairment: findings from the Weiss Functional Impairment Rating Scale. *Evidence-Based Mental Health*, *21*, 155-164. http://dx.doi.org/10.1136/ebmental-2018-300025

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K. et al. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. *Archives of General Psychiatry*, *69*, 274-282. https://doi.org/10.1001/archgenpsychiatry.2011.147

Whiteside, S. P. (2009). Adapting the Sheehan disability scale to assess child and parent impairment related to childhood anxiety disorders. *Journal of Clinical Child & Adolescent Psychology*, *38*, 721-730. https://doi.org/10.1080/15374410903103551

Youngstrom, E. A., Salcedo, S., Frazier, T. w., & Perez Algorta, G. (2019). Is the finding too good to be true? Moving from "more is better" to thinking in terms of simple predictions and credibility. *Journal of Clinical Child and Adolescent Psychology, 48,* 811-824. https://doi.org/10.1080/15374416.2019.1669158

Zahra, D., Qureshi, A., Henley, W., Taylor, R., Quinn, C., Pooler, J. et al. (2014). The Work and Social Adjustment Scale: Reliability, sensitivity and value. *International Journal of Psychiatry in Clinical Practice*, *18*, 131-138. https://doi.org/10.3109/13651501.2014.894072

Table 1

Study 2: Means (M), Standard Deviations (SD), and Internal Consistency Estimates (α) of Survey Measures

Variable	*M*	*SD*	α
Work and Social Adjustment Scale (WSASY Total Scores)			
Youth [a]	7.29	5.46	.74
Youth Square-Root Transformed Score	2.46	1.12	—
Teacher [b]	3.98	5.58	.86
Teacher Square-Root Transformed Score	1.37	1.45	—
Parent [c]	6.22	6.68	.83
Brief Problem Monitor (BPM Mean Raw Scores)			
Youth	0.48	0.29	.84
Teacher	0.17	0.24	.88
Teacher Square-Root Transformed Score	0.30	0.29	—
Parent	0.30	0.26	.83
Parent Square-Root Transformed Score	0.50	0.24	—
Social Skills Improvement System-Rating Scales (SSIS-RS Mean Raw Scores of 21-Item Version)			
Youth	2.21	0.38	.86
Teacher	2.63	0.46	.95
Parent	2.32	0.42	.91

Notes. [a] Estimates for youth reports based on 66 WSASY reports, 65 BPM reports, and 65 SSIS-RS reports. [b] Estimates for teachers' reports based on 63 WSASY reports, 64 BPM reports, and 61 SSIS-RS reports. [c] Estimates for parents' reports based on 66 WSASY reports, 62 BPM reports, and 65 SSIS-RS reports.

Table 2

Study 2: Bivariate Correlations among Multi-Informant Reports on the Work and Social Adjustment Scales for Youth (WSASY) and Continuous Criterion Measures (i.e., Classroom Observations, Brief Problem Monitor [BPM], and Social Skills Improvement System-Rating Scales [SSIS-RS])

Variable	**Classroom Observations of Youth Psychosocial Difficulties** [b]	**BPM (Youth)**[c]	**BPM (Teacher)**[c]	**BPM (Parent)**[c]	**SSIS-RS (Youth)**[d]	**SSIS-RS (Teacher)** [d]	**SSIS-RS (Parent)**[d]
WSASY							
Youth	.12	.52***	.26*	.37**	-.58***	-.51***	-.37**
Teacher	.51***	.53***	.87***	.46***	-.41**	-.76***	-.37**
Parent	.12	-.07	.02	.63***	.09	-.24	-.57***
Simple Average of Cross-Informant Correlations [a]	.25	.32	.38	.48	-.30	-.50	-.43
Trait Score	.36**	.55***	.60***	.65***	-.51***	-.75***	-.61***

Note. [a] To compare against the magnitude of correlations observed with the WSASY *Trait* score, we computed a simple average of the three correlations reported in the same column based on the individual WSASY youth, teacher, and parent reports. [b] Correlations for classroom observations based on data about 57 youth for WSASY youth reports, 57 WSASY teacher reports, and 57 WSASY parent reports. [c] Correlations for Brief Problem Monitor based on data from 59-63 WSASY youth reports, 59-63 WSASY teacher reports, and 59-63 WSASY parent reports, depending on pairwise Brief Problem Monitor data. [d] Correlations for Social Skills Improvement System-Rating Scales based on data from 58-62 WSASY youth reports, 58-62 WSASY teacher reports, and 58-62 WSASY parent reports, depending on pairwise Social Skills Improvement System-Rating Scales data. *$p < .05$; **$p < .01$; ***$p < .001$.

Table 3

Study 2: Independent Samples t-Tests Examining Multi-Informant Reports on the Work and Social Adjustment Scales for Youth (WSASY) and Discrete Measure Based on Home Observations of Youth Psychosocial Difficulties

Group	**WSASY Youth *M*(SD)**	**WSASY Teacher *M*(SD)**	**WSASY Parent *M*(SD)**	**WSASY *Trait* Score *M(SD)***	***t*-Test(df)**	***p*-Value**	**Cohen's *d***
Score above "0" (n = 16)	2.78 (1.11)	1.98 (1.68)	9.18 (7.56)	0.53 (1.10)	**Youth WSASY:** 1.28(61)	.20	0.37
					Teacher WSASY: 1.95(61)	.055	0.53
					Parent WSASY: 2.24(61)	.03	0.61
Score of "0" (n = 47)	2.37 (1.09)	1.17 (1.33)	5.02 (6.01)	-0.18 (0.90)	Simple Average of Cross-Informant Effect Sizes [a]	—	0.50
					***Trait* Score WSASY:** 2.60(61)	.01	0.70

Note. [a] To compare against the magnitude of effect sizes observed with the WSASY *Trait* score, we computed a simple average of the three effect sizes reported above based on the individual WSASY youth, teacher, and parent reports.

The Work and Social Adjustment Scale for Youth: Psychometric Properties of the Teacher Version and Evidence of Contextual Variability in Psychosocial Impairments

ONLINE SUPPLEMENTARY MATERIAL

Table S1

Study 2: Principal-Component Analysis (PCA) of Multi-Informant Reports on the Work and Social Adjustment Scale for Youth (WSASY)

Informant	Trait	Context	Perspective	Sources of Variability in Informant's Report
WSASY (*n* = 63)	**Component Weight**			
Youth	0.81	-0.03	-0.58	Home and School (context), Self (perspective)
Teacher	0.69	-0.53	0.48	School (context), Other (perspective)
Parent	0.48	0.82	0.29	Home (context), Other (perspective)
Eigenvalue	1.37	0.97	0.66	
Variance attributable to each component	45.60%	32.46%	21.93%	

Note. PCA conducted with participants for whom we had full data across informants' WSASY reports. Kraemer et al. (2003) described the following criteria for each component: *Trait* (all three informants' reports load strongly and in the same direction), *Context* (informants from different contexts load in opposite directions) and *Perspective* (self-reports load in the opposite direction of observer informants' reports).

Work and Social Adjustment Scale for Youth: Teacher Version

The ways children and adolescents think, feel, or behave sometimes affect their ability to do everyday things. These things might include doing well in school, completing household chores, relaxing during free time, and having close relationships with friends and family. We would like you to look at each of the items below and rate **THIS STUDENT** on how much the ways s/he thinks, feels, or behaves impair her/his ability to do each of the everyday things described in the items. By "impair" we mean "make difficult, harm, or worsen." An example might be if the ways your student thinks, feels, or behaves create problems for her/him that get in the way of completing homework assignments or making friends.

On a scale of 0 - 8, where 0 means 'not at all impaired' and 8 means 'very severely impaired', please rate the following:

1. Because of the ways **this student** thinks, feels, or behaves, her/his **ability to do well in school** is impaired.

Not at all		Slightly		Definitely		Severely		Very Severely
0	1	2	3	4	5	6	7	8

2. Because of the ways **this student** thinks, feels, or behaves, her/his **ability to complete class assignments** is impaired**.**

Not at all		Slightly		Definitely		Severely		Very Severely
0	1	2	3	4	5	6	7	8

3. Because of the ways **this student** thinks, feels, or behaves, her/his **ability to enjoy free time spent *with other people*** is impaired**.** *(For example, socializing during or between classes or at lunch, participating in school social activities, clubs, or sports).*

Not at all		Slightly		Definitely		Severely		Very Severely
0	1	2	3	4	5	6	7	8

4. Because of the ways **this student** thinks, feels, or behaves, her/his **ability to enjoy free time spent *alone*** is impaired**.** *(For example, reading, hobbies, listening to or playing music, or working independently).*

Not at all		Slightly		Definitely		Severely		Very Severely
0	1	2	3	4	5	6	7	8

5. Because of the ways **this student** thinks, feels, or behaves, her/his **ability to form and maintain close relationships with other people** is impaired. *(For example, classmates, friends, teachers, or other school staff.)*

Not at all		Slightly		Definitely		Severely		Very Severely
0	1	2	3	4	5	6	7	8

Note. Adapted with permission from "The Work and Social Adjustment Scale for Youth: A measure for assessing youth psychosocial impairment regardless of mental health status," by A. De Los Reyes et al., 2019, *Journal of Child and Family Studies*, *28*(1), pp. 13-14 (https://doi.org/10.1007/s10826-018-1238-6). Copyright 2017 by Springer.