



Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Test use and assessment practices of school psychologists in the United States: Findings from the 2017 National Survey[☆]

Nicholas F. Benson^{a,*}, Randy G. Floyd^b, John H. Kranzler^c, Tanya L. Eckert^d, Sarah A. Fefer^e, Grant B. Morgan^a

^a Baylor University, United States of America

^b University of Memphis, United States of America

^c University of Florida, United States of America

^d Syracuse University, United States of America

^e University of Massachusetts, United States of America



ARTICLE INFO

Associate Editor: Stephen Kilgus

Keywords:

Assessment
Psychological testing
Classification
Diagnostic practices

ABSTRACT

Although several studies have examined the specific instruments and procedures used by school psychologists when conducting comprehensive psychoeducational evaluations, the last one was published over 20 years ago (viz., Wilson & Reschly, 1996). Given the substantial theoretical and practical advances in assessment since then, the purpose of the current study was to examine the test use and assessment practices of contemporary school psychologists in the United States. Data from the 2017 National Survey of Assessment Practices in School Psychology revealed that test use and assessment practices have evolved significantly. Much of this change consists of the substitution of tests and practices with limited reliability and validity with those with greater psychometric support. Results of this study also indicate that school psychologists regularly conduct multi-method assessments to prevent, identify, monitor, and remediate child and adolescent learning difficulties and other presenting problems in the schools.

1. Introduction

In the 2014–2015 academic year, 13% of all children and youth enrolled in the public school system – or 6.6 million students – received special education and related services (National Center for Education Statistics, 2018). Since the enactment of the Individuals with Disabilities Education Act (IDEA, 2004), formerly known as the Education for All Handicapped Children Act, school psychologists have served as an integral member of the multidisciplinary team that determines eligibility for special education and related services because they are uniquely qualified to conduct the requisite comprehensive psychoeducational evaluations for eligibility determination. As stated in the *Model for Comprehensive and Integrated School Psychological Services* (National Association of School Psychologists [NASP], 2010), “school psychologists have knowledge of varied models and methods of assessment and data

[☆] The authors wish to extend our gratitude to the many state associations who promoted our survey to members. We also thank NASP for allowing us to recruit randomly selected members. We extend our appreciation to the following scholars for their reviews and comments on the survey: Craig Albers, Jamilia Blake, Amy Briesch, Bridget Dever, Scott Graves, Ryan Kettler, Steve Kilgus, Ben Lovett, Ryan McGill, Kara McGoey, Chris Niileksela, Matt Reynolds, Chris Riley-Tillman, Don Saklofske, and Sara Whitcomb. Finally, we recognize Carmen McGuinness for her assisting in preparing this manuscript.

* Corresponding author at: Department of Educational Psychology, Baylor University, One Bear Place #97301, Waco, TX 76798, United States.
E-mail address: Nicholas_Benson@baylor.edu (N.F. Benson).

<https://doi.org/10.1016/j.jsp.2018.12.004>

Received 31 May 2018; Received in revised form 18 September 2018; Accepted 19 December 2018

Available online 30 December 2018

0022-4405/ © 2019 Society for the Study of School Psychology. Published by Elsevier Ltd. All rights reserved.

collection for identifying strengths and needs, developing effective services and programs, and measuring progress and outcomes” (p. 4). Although school psychologists provide a comprehensive array of psychological services in the schools, assessment is central to the data-based decision-making and accountability practices that permeate all aspects of service delivery (NASP, 2010).

Over the past 30 years, most estimates suggest that school psychologists spend about 50% of their time in assessment activities related to eligibility determination (e.g., Goh, Teslow, & Fuller, 1981; Hutton, Dubes, & Muir, 1992; Stinnett, Harvey, & Oehler-Stinnett, 1994; Walcott & Hyson, 2018). For example, in a recent survey of NASP members, Walcott and Hyson (2018) asked school psychologists to rate the frequency with which they participated in different psychological and educational practices. Frequency ratings were based on a scale with values ranging from 0 to 4, with 0 indicating “No Involvement” in the activity and 4 indicating a “Great Deal” of involvement. School psychologists reported that the two professional practices they engaged in most frequently were conducting psychoeducational evaluations to determine students’ eligibility for special education services ($M = 3.31$) and developing IEPs ($M = 2.81$). Of course, school psychologists engage in many other assessment activities to inform effective decision making and problem solving in the schools, such as academic and behavioral progress monitoring, as well as universal screening for mental health services (e.g., Splett, Fowler, Weist, McDaniel, & Dvorsky, 2013). These assessment activities can be used to independently to identify which students require school-based interventions or evaluate the effectiveness of these programs. These assessment activities can also be incorporated as part of comprehensive psychoeducational evaluations.

1.1. Assessment in school psychology

Comprehensive psychoeducational evaluations consist of a variety of instruments and procedures and can include gathering information from different informants, in different settings, and using a variety of procedures, such as interviews, observations, rating scales and checklists, and standardized and non-standardized tests. Which psychological instruments and practices do school psychologists use when conducting these assessments? A number of studies have addressed this question (Goh et al., 1981; Hutton et al., 1992; Reschly, Genshaft, & Binder, 1987; Stinnett et al., 1994; Wilson & Reschly, 1996).

1.1.1. Goh et al. (1981)

In the first such study, Goh et al. (1981) surveyed 274 members of NASP on their use of tests in the following eight categories: intelligence, achievement-general, achievement-specific, perceptual-motor functioning, personality, behavior rating scales, preschool evaluation, and vocational interest. Blank spaces were provided so that respondents could indicate their use of instruments not listed. Goh et al. did not ask respondents to report about more informal procedures, such as unstructured interviewing and only inquired about informal observation in the category of early childhood/developmental assessment. For each test listed, they asked respondents to report the frequency of usage on a scale ranging from 0 to 3, with 0 indicating “No Use” and 3 indicating “Frequent Use.” With these data, they created three scores: the total mention score (TMS) was total number of times respondents indicated that they used each of the instruments listed; the usage rating score (URS) was the sum of all usage ratings for each instrument; and the weighted usage score (WUS) was calculated by multiplying each respondent’s usage ratings for an instrument “by the percentage of his or her total assessment devoted to that particular assessment area” (p. 701).

Results of their survey revealed that tests of intelligence (27%), perceptual and motor functioning (22%), achievement (22%), and personality (14%) accounted for 85% of all instruments administered (Goh et al., 1981). The personality tests that were given tended to consist of quick and easy-to-use projective tests (e.g., sentence completion and drawing tests) rather than more time-consuming techniques, such as the Thematic Apperception Test (Morgan & Murray, 1935). Behavior rating scales, which included measures of adaptive behavior, accounted for only 9% of all tests used, with informal procedures used more frequently than published rating scales. In addition, preschool evaluations and vocational interest assessment were conducted by a very small percentage of practitioners. These results suggested that, at that time, a typical battery administered in comprehensive evaluations for eligibility determination consisted of tests of intelligence, achievement, perceptual and motor functioning, and personality.

Table 1 presents a Top 10 list of the assessment instruments used most by school psychologists across surveys. Please note that the instruments listed in this table reflect all versions of each test published over the years in which these studies were conducted. For example, the ranking for the Wechsler Intelligence Scales across studies is based on the reported use of the Wechsler Intelligence Scales for Children, Revised (Wechsler, 1974) and the Wechsler Intelligence Scales for Children, Third Edition (Wechsler, 1991). It is also important to note that, although Goh et al. (1981) listed the TMS for each instrument, they reported only the rank order of instruments within each category for the URS and WUS scores, not the scores themselves. The only overall Top 10 ranking that can be made for Goh et al.’s survey, therefore, is based on the TMS. The TMS reflects how widespread and not how often an instrument is used. As can be seen in this table, in the early 1980s, the Wechsler Intelligence Scales were the most widely used tests. In fact, 4 of the Top 10 most widely used instruments were intelligence tests: the Wechsler scales, the Stanford-Binet (Thorndike, Hagen, & Sattler, 1986), the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997), and the Goodenough-Harris Draw-A-Man Test (Harris, 1963). The remaining tests in the Top 10 list consisted of two standardized achievement tests (i.e., Wide Range Achievement Test [WRAT], Wilkinson, 1993; Peabody Individual Achievement Test [PIAT], Markwardt, 1989); two projective tests (i.e., sentence completion test and the Bender Test of Visual-Motor Gestalt [Bender-Gestalt], Bender, 1946); and one visual-motor integration test. As is evident in these results, the Bender-Gestalt was widely used both as a projective test and visual-motor integration test.

1.1.2. Reschly et al. (1987)

In the second national survey conducted, Reschly et al. (1987) gathered data from 242 members of NASP on their test usage and practices in the following nine categories: intelligence, adaptive behavior, achievement-screening, achievement-diagnostic,

Table 1

Top 10 list of instruments and procedures used most by school psychology practitioners across surveys.

	Goh et al. (1981) ^a	Reschly et al. (1987) ^b	Hutton et al. (1992) ^a	Wilson and Reschly (1996) ^{b,c}	Stinnett et al. (1994) ^{d,e}
Instrument/procedure					
Wechsler Intelligence Scales	1	1	1	2	2
Wide Range Achievement Tests	2	6	3	10	10
Stanford-Binet Intelligence Tests	3	–	–	–	–
Bender Gestalt (visual motor)	4	3	–	3	4
Peabody Picture Vocabulary Test	5	–	10	–	–
Bender Gestalt (personality)	6	–	–	–	6
Goodenough-Harris Draw-A-Man Test	7	–	8.5	–	7
Sentence Completion Test	8	–	–	–	–
Adaptive Behavior Scales (non-specific)	9	–	5	–	–
Peabody Individual Achievement Test	10	–	4	–	–
Unstructured Interview (Teacher/Child/ Parent)	–	2	–	–	1
Draw-A-Person	–	4	–	4	9
Anecdotal Observation	–	5	–	5	3
Structured Observations	–	7	–	1	–
Woodcock-Johnson Tests of Achievement	–	8	–	9	5
House-Tree-Person	–	9	–	6	–
Kinetic Figure Drawing	–	10	–	8	–
Visual-Motor Integration	–	–	–	7	–
KeyMath Diagnostic Assessment	–	–	2	–	–
Vineland Adaptive Behavior Scales	–	–	6	–	–
Stanford Achievement Test	–	–	7	–	–
Illinois Test of Psycholinguistic Abilities	–	–	8.5	–	–
Curriculum-Based Assessment	–	–	–	–	8

Note.^a Ranking based on total mention score.^b Ranking based on number of tests and procedures conducted per month.^c Data for this study were gathered in 1991–1992.^d Data for this study were gathered in 1992.^e Ranking based on number of tests and procedures conducted per school year.

perceptual/motor, projectives/figure drawing, informal instruments/procedures, checklists/rating scales, and batteries and miscellaneous. In contrast to Goh et al. (1981), Reschly et al. asked respondents to indicate the number of times they used each instrument or procedure per month. Thus, although also based on self-report, their results provide a more direct assessment of the frequency with which school psychologists used particular instruments and procedures. In addition, they added an informal instruments/procedures category to assess the number of times practitioners conducted structured and unstructured interviews and classroom observations. Reschly et al. did not ask respondents to report test usage in the categories of preschool evaluation and vocational interest.

Despite the differences between these two surveys, Reschly et al.'s (1987) results were quite similar to those of Goh et al. (1981). Reschly et al. found that a typical comprehensive evaluation consisted of tests of intelligence, achievement, perceptual/motor functioning, and personality (projectives/figure drawings), but they also found that school psychologists regularly conducted structured and unstructured interviews and classroom observations. The use of checklists/rating scales was low, in contrast. Results of their survey indicated that tests of intelligence (27%), perceptual/motor functioning (22%), achievement (22%), and personality (14%) accounted for 85% of all instruments administered. Behavior rating scales, which included measures of adaptive behavior, accounted for only 9% of all tests used, with informal procedures used more frequently than published rating scales. Early childhood/developmental assessment and vocational interest assessment were conducted by a small percentage of practitioners.

As can be seen in Table 1, Reschly et al.'s results indicated that, in the mid-1980s, not only were the Wechsler Intelligence Scales the most widely used instruments by school psychologists, they were also the most frequently administered. At that time, respondents indicated that they administered the Wechsler scales an average of 9.92 times per month ($SD = 7.16$). The Wechsler scales are the only intelligence tests listed in the Top 10 most used instruments. The large standard deviation reflects a high degree of variability in the use of these scales, however. The second most used test was the PPVT-Revised (Dunn & Dunn, 1997) which, in contrast to the Goh et al. (1981) study, was not classified as an intelligence test. The PPVT-Revised was administered 2.42 times per month ($SD = 5.01$). In addition, these rankings showed that school psychologists frequently conducted unstructured interviews with teachers, children, and, to a lesser extent, parents. Reschly et al. also found that respondents regularly engaged in anecdotal and structured observations. The third most widely used instrument was the Bender-Gestalt to assess visual-motor integration (not as a projective test). In addition, they found that three of the most frequently used instruments were projective tests. The personality tests administered most often consisted of quick and easy-to-use projective drawings, such as the Draw-A-Person, the House-Tree-Person (e.g., Buck, 1964), and the Kinetic Family Drawing. Two standardized tests of general achievement were also listed in the Top 10: the WRAT (Wilkinson, 1993) and the Woodcock-Johnson (WJ) Tests of Achievement (Woodcock & Johnson, 1989).

1.1.3. Hutton et al. (1992)

In a study intended to update the results of Goh et al. (1981) a decade later, Hutton et al. (1992) surveyed 389 members of NASP on their use of psychological tests. They used the same categories as Goh et al. but separated behavior rating scales into categories for behavior ratings and adaptive behavior. Hutton et al. did not assess the use of structured and unstructured interviews or classroom observations. Given that some tests go out of print, others are revised, and new tests are introduced over time, the specific tests listed differed somewhat from Goh et al. Using the same format as the earlier study, Hutton et al. asked respondents to indicate the frequency of usage for each instrument on a scale ranging from 0 to 3, with 0 indicating “No Use” and 3 indicating “Frequent Use.” Results of Hutton et al.’s survey indicated some noteworthy changes in the test usage of school psychologists 10 years later. Their findings revealed a decline in the use of perceptual and motor functioning tests and an increase in the administration of achievement tests, adaptive behavior measures, and behavior rating scales. Their results suggested that the typical comprehensive test battery administered by school psychologists consisted of tests of intelligence, achievement, personality, and behavioral rating and adaptive behavior measures.

Because Hutton et al. (1992) reported only the TMS score for each instrument and rank ordered instruments based on that score within each category, Table 1 displays the ranking of the most used instruments and procedures in their survey. As can be seen here, Hutton et al.’s results show that, in the early 1990s, the Wechsler Intelligence Scales remained the most widely used tests by school psychologists. However, only 2 of the Top 10 most widely used instruments were intelligence tests: the Wechsler scales and the Draw-A-Man test. They also found that 5 of the Top 10 most widely used instruments were general and specific standardized tests of achievement: the WRAT, the PIAT, the Stanford Achievement Test (The Psychological Corporation, 1988), the Metropolitan Achievement Test (Balow, Farr, & Hogan, 1992), and the Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy, & Kirk, 1968). As reflected in the ranking of the Vineland Adaptive Behavior Scale (Sparrow, Balla, & Cicchetti, 1984), school psychologists were using adaptive behavior scales more widely than a decade before. No tests of personality or visual-motor integration were listed in the Top 10.

1.1.4. Wilson and Reschly (1996)

Five years later, Wilson and Reschly (1996) conducted a replication and extension of Reschly et al.’s (1987) survey. They gathered data in 1991–1992 from 1089 practitioners who were members of NASP on their test usage and practices in seven categories: ability/intelligence; preschool and family; social/emotional, adaptive behavior, and social skills; educational skills; visual/motor; projective personality techniques; and systematic behavior observation. They asked respondents to indicate the number of times they administered each instrument or procedure per month. Results of their study revealed few differences in test use and assessment practices from the earlier survey. The one exception was the use of structured observations, which Wilson and Reschly found were the most used assessment procedure at that time. As they noted, however, this might have been due to differences in how the two surveys asked respondents to report about observation. Reschly et al. asked respondents to report on the number of times they conducted structured and unstructured observations, whereas Wilson and Reschly asked respondents to report on the number of times they conducted four different kinds of observation (i.e., event, duration, time sampling, and interval). As a result, this may have resulted in duplication in reporting and therefore higher usage results. In any case, although one other test was used by a higher percentage of school psychologists, the Wechsler Intelligence Scales were easily the instruments used most often. The Bender-Gestalt (not personality) was the third most widely used instrument. The listing of the Visual-Motor Integration test in the Top 10 suggested an increased emphasis on the assessment of visual-motor assessment since Reschly et al.’s (1987) survey. The WJ Tests of Achievement and the WRAT were the only standardized tests of achievement listed. Interestingly, unstructured interviews (teacher/child/parent) were not listed in the Top 10. Last, as the authors noted, behavior rating and adaptive behavior measures were not regularly used at that time.

1.1.5. Stinnett et al. (1994)

Stinnett et al. (1994) surveyed 123 school-based practitioners in May of 1992 who were members of NASP. Although published before Wilson and Reschly’s survey, it reflects the most recent data on the test usage and assessment practices of school psychologists from over 20 years ago. They asked respondents to indicate the number of times they used the following instruments or procedures per school year listed in the following seven categories: intelligence/cognitive, academic achievement, perceptual and perceptual motor, behavioral/social/emotional, adaptive behavior, vocational interests and aptitude, and preschool functioning. They also asked respondents to report the perceived importance of the results of each test or procedure on a 3-point Likert scale, ranging from “Not Important” to “Very Important.”

Results of Stinnett et al. were similar to the results of Reschly et al. (1987) and Wilson and Reschly (1996) insofar as respondents reported frequent use of tests of intelligence, academic achievement, behavioral/social/emotional functioning, and perceptual motor functioning. In addition, they also found that the assessment of vocational interests and preschool functioning was conducted by relatively few school psychologists. Results of Stinnett et al. (1994), however, revealed that the number of specific instruments and procedures administered each year varied greatly. As they stated, this “suggests that a broader repertoire of instruments and procedures is being used, and except for the major instruments, also suggests a lack of consensus by practitioners as to what are the best tools and practices in assessment” (p. 342). For example, although only 23% of respondents reported use of curriculum-based assessment (CBA), those who reported using it did so extensively. In terms of the importance of the instruments and procedures used, standardized norm-referenced tests of intelligence and achievement were rated as highly important, as were structured and unstructured interviews and observations. In contrast, although tests of perceptual motor functioning continued to be administered frequently, they were rated as relatively less important. In general, instruments and procedures rated as the least important tended to

be those with questionable psychometric properties, such as projective techniques.

The Top 10 most used instruments and procedures in Stinnett et al.'s (1994) survey are shown in Table 1. As can be seen here, their Top 10 ranking overlaps to some degree with those of Reschly et al. (1987) and Wilson and Reschly (1996), but findings also differed. The most frequently used tests or procedures were unstructured interviews (administered to teachers, children, and parents), followed by the Wechsler scales. Anecdotal observations were ranked fourth. Despite the fact that structured observations were not ranked in the Top 10, they were the second most frequently used measure in the behavioral/social/emotional category. The Bender-Gestalt continued to be the most frequently administered test of visual-motor integration. It is noteworthy that, in addition to the two standardized tests of achievement listed in the Top 10 (WJ Tests of Achievement and WRAT), CBA was listed for the first time, reflecting the increased emphasis on the direct assessment of academic functioning that began in the early 1990s. Last, although the House-Tree-Person and Kinetic Family Drawing were not listed in the Top 10, the frequent use of the Draw-A-Person and the Bender-Gestalt indicated that school psychologists regularly administered projective tests. One likely explanation for the difference in results from the Reschly et al. and Wilson and Reschly surveys is that the Bender-Gestalt was listed as a test of behavioral/social/emotional functioning in the Stinnett et al. study and only listed in the perceptual-motor categories in the other two surveys.

1.1.6. Summary

Taken as a whole, results of these surveys indicated that the comprehensive evaluations conducted by school psychologists in the 1980s and early 1990s consisted largely of tests of intelligence, academic achievement, perceptual and motor functioning, and personality, although structured and unstructured interviews and classroom observations were also widely conducted. Across studies, the Wechsler Intelligence Scales were consistently the most frequently used instruments, followed by the Bender-Gestalt and the Draw-A-Person. The two most widely used tests of academic achievement were the WRAT and the WJ Tests of Achievement, and CBA was listed for the first time among the Top 10 most widely used instruments in the most recent survey by Stinnett et al. (1994). The personality tests that were administered most tended to be projective drawings. Overall, results of these surveys were largely encouraging, mainly because they showed that school psychologists tended to conduct comprehensive, multi-method assessments with instruments that were generally sound psychometrically. Both Wilson and Reschly (1996) and Stinnett et al. (1994), however, were critical of the continued use of some instruments with poor psychometric properties, particularly projective drawings.

1.2. Purpose of the current study

Although school psychologists provide a broad array of psychological services in the schools, assessment has long been a prominent role (e.g., Walcott & Hyson, 2018). The last comprehensive survey of school psychologist test usage and assessment procedures, however, was published over 20 years ago. Much has changed since then, including revisions to special education legislation, increased variation in assessment methods, and advances in administration and scoring technology. For example, when the IDEA was re-authorized in 2004, the requirement of a “severe discrepancy” between scores on cognitive ability and achievement tests was eliminated for identification of specific learning disabilities (SLD). To encourage the use of procedures to identify SLD that are more relevant to classroom instruction, a provision was added allowing the use of response-to-intervention (RtI) methods, which refers to “a process that determines if the child responds to scientific, research-based intervention as a part of the evaluation procedures” (IDEA, 2004, see Section 614[b][6]), when determining SLD. As an extension of RtI methods, multitiered systems of support (MTSS) have been widely used not only for the delivery of psychological services in the schools, but also as an alternative to the intelligence-achievement discrepancy method for the identification of SLD (e.g., Jimerson, Burns, & VanDerHeyden, 2016). The re-authorization of IDEA, therefore, may have led to significant changes in test use and procedures since the last comprehensive assessment survey of school psychologists was conducted by Stinnett et al. (1994), particularly concerning the use of intelligence tests.

In addition, significant advances in research and theory over the past two decades have resulted in the introduction of new methods of assessment, revision of extant tests, and the development of new instruments. For example, since the 1990s, the Cattell-Horn-Carroll (CHC) theory of the structure of cognitive abilities (see Schneider & McGrew, 2018) has been increasingly used by test developers and practitioners to guide the assessment of cognitive abilities (e.g., Sotelo-Dynega & Dixon, 2014). At this time, most contemporary tests of intelligence are either explicitly based on CHC theory or have been heavily influenced by it (e.g., Kranzler, Benson, & Floyd, 2016). In addition, behavior rating scale technology has improved significantly over this time, as is reflected in the development and publication of a number of broad-band instruments measuring child and adolescent behavior problems (e.g., externalizing and internalizing behaviors), as well as numerous narrow-band instruments targeting specific indicators of child psychopathology (e.g., ADHD; Whitcomb, 2017). Rating scales also have been designed to assess both adaptive and problem behaviors, such as the Behavior Assessment System for Children, Third Edition (BASC-3; Reynolds & Kamphaus, 2015a).

Moreover, over the past 20 years the United States has become increasingly culturally and linguistically diverse. Consequently, new tests and procedures have been developed to assess this population of diverse learners, such as multidimensional nonverbal tests of intelligence (e.g., Universal Nonverbal Intelligence Test, Second Edition [UNIT2]; Bracken & McCallum, 2016). In addition, the widespread adoption of MTSS in the schools, which requires on-going screening and progress monitoring, has likely impacted use of more direct and frequent assessments of academic achievement and social behavior (Shapiro, Angello, & Eckert, 2004). The use of MTSS may have also resulted in changes in the use of individually-administered standardized tests of achievement. Also, as indicated by the prevalence rate for ASD increasing faster than that of any other neurodevelopmental disorder, there has been an increased need for school-based professionals to identify students with ASD using evidence-based measures (Wilkinson, 2010). Last, when Wilson and Reschly (1996) published the last of these early surveys, the world-wide web had just been invented. Use of the internet has increased exponentially since that time, as have advances in computing technology, including computer-adaptive testing. As a

Table 2
Reported racial and ethnic characteristics of the sample.

Characteristic	Frequency	Percent
Race		
American Indian or Alaskan Native	5	0.4
Asian	19	1.4
Biracial	11	0.8
Black or African American	33	2.5
Middle Eastern/North African	3	0.2
Multiracial	16	1.2
Native Hawaiian or Pacific Islander	1	0.1
White	1203	91.5
Prefer not to say	24	1.8
Ethnicity		
Hispanic or Latino	79	6.1
Not Hispanic or Latino	1199	91.9
Prefer not to say	26	2.0

result, the impacts of these developments on the test usage and assessment practices of school psychologists are largely unknown at present.

Given that the last comprehensive survey of the test usage and assessment procedures of school psychologists was published over two decades ago, the purpose of the current study was to examine the test use and assessment practices of contemporary school psychologists across the nation. Results of this survey will provide current information regarding the most widely used assessment instruments and practices. Furthermore, these findings will have implications for faculty in school psychology training programs, informing the content and scope of their assessment curriculum for the optimal preparation of future school psychologists. Finally, these results will help guide future research regarding assessment practices.

2. Method

This research project was determined to be exempt from review by the Baylor University Institutional Review Board according to federal regulation 45 CFR 46.101(b)(2). Participation was voluntary, and all participants completed an electronic consent waiver. In the sections that follow, we report survey methods according to the Checklist for Reporting Results of Internet ESurveys (Eysenbach, 2004).

2.1. Participants

Participants were 1317 school psychologists who were practicing in the United States. Detailed racial and ethnic information is presented in Table 2. About 83% of the participants were exclusively licensed or certified to work in school settings, and roughly 17% held additional licensures that allow them to practice outside of school settings (e.g., licensed psychologists). Approximately 50% of the sample reported being a Nationally Certified School Psychologist (NCSP). The average years of experience as a practicing school psychologist, excluding internship, was 10 ($SD = 8.6$); the mean age was 39 ($SD = 11.0$), which is 7% lower than the average of 42 found in the most recent NASP Membership Survey (Walcott & Hyson, 2018). Most of the sample reported being women (90%), which is 8% higher than found in the NASP Membership Survey. About 91% of the sample reported being White, which is 6% higher. The highest degree earned by most of the sample was a specialist degree (65%), which is 15% higher. In contrast, only 14% percent reported earning a doctorate relative to 25% in the NASP Membership Survey. About 84% of the sample reported graduating from a program that was approved by NASP, accredited by the American Psychological Association (APA), or both. Overall, the demographic characteristics of participants in this study are quite comparable to those of the most recent survey of NASP members (Walcott & Hyson, 2018).

Frequency counts of participants by state are presented in Table 3. As can be seen here, the sample is reasonably representative of the U.S. population across the 50 states. Additionally, we compared the percentage of the U.S. population with the percentage of participants in the survey by state. All states were well represented in the survey relative to percent of U.S. population. Although California was the most underrepresented state, it also had the highest number of participants in the survey.

2.2. Recruitment of participants

Multiple recruitment strategies were employed to obtain a large and representative sample. All methods featured a recruitment letter that included an anonymous link to the survey. Recruitment also involved informing potential participants that they had the option to participate in a drawing for one of six \$50 gift certificates. A link was provided at the end of the survey so that contact information was collected separately from survey responses.

First, we submitted a proposal to NASP, who approved the proposal and provided us with a list of member mailing addresses. Per NASP policy we were limited to a list of addresses for 1000 randomly selected members and email addresses were not provided. We

Table 3
Representativeness of the sample by state.

State	Frequency	Survey %	Population %	Survey % - population %
Alabama	8	0.60%	1.51%	-0.91%
Alaska	3	0.20%	0.23%	-0.03%
Arizona	18	1.40%	2.15%	-0.75%
Arkansas	6	0.50%	0.93%	-0.43%
California	128	9.70%	12.15%	-2.45%
Colorado	21	1.60%	1.72%	-0.12%
Connecticut	35	2.70%	1.11%	1.59%
Delaware	9	0.70%	0.30%	0.40%
Florida	60	4.80%	6.38%	-1.58%
Georgia	36	2.70%	3.19%	-0.49%
Hawaii	6	0.50%	0.44%	0.06%
Idaho	4	0.30%	0.52%	-0.22%
Illinois	55	4.20%	3.96%	0.24%
Indiana	34	2.60%	2.05%	0.55%
Iowa	55	4.20%	0.97%	3.23%
Kansas	18	1.40%	0.90%	0.50%
Kentucky	14	1.10%	1.37%	-0.27%
Louisiana	14	1.10%	1.45%	-0.35%
Maine	4	0.30%	0.41%	-0.11%
Maryland	24	1.80%	1.86%	-0.06%
Massachusetts	44	3.30%	2.11%	1.19%
Michigan	19	1.40%	3.07%	-1.67%
Minnesota	17	1.30%	1.71%	-0.41%
Mississippi	3	0.20%	0.93%	-0.73%
Missouri	11	0.80%	1.89%	-1.09%
Montana	7	0.50%	0.32%	0.18%
Nebraska	37	2.80%	0.59%	2.21%
Nevada	10	0.80%	0.91%	-0.11%
New Hampshire	6	0.50%	0.41%	0.09%
New Jersey	59	4.50%	2.77%	1.73%
New Mexico	2	0.20%	0.64%	-0.44%
New York	97	7.40%	6.11%	1.29%
North Carolina	20	1.50%	3.14%	-1.64%
North Dakota	6	0.50%	0.24%	0.26%
Ohio	79	6.00%	3.59%	2.41%
Oklahoma	9	0.70%	1.21%	-0.51%
Oregon	7	0.50%	1.27%	-0.77%
Pennsylvania	60	4.60%	3.96%	0.64%
Rhode Island	5	0.40%	0.33%	0.07%
South Carolina	14	1.10%	1.54%	-0.44%
South Dakota	15	1.10%	0.27%	0.83%
Tennessee	15	1.10%	2.06%	-0.96%
Texas	110	8.40%	8.62%	-0.22%
Utah	10	0.80%	0.94%	-0.14%
Vermont	9	0.70%	0.19%	0.51%
Virginia	23	1.70%	2.60%	-0.90%
Washington	32	2.40%	2.26%	0.14%
West Virginia	6	0.50%	0.57%	-0.07%
Wisconsin	24	1.80%	1.79%	0.01%
Wyoming	8	0.60%	0.18%	0.42%

mailed these members a hard copy of the recruitment letter during June, 2017 that included an anonymous link to the survey. As responses were anonymous, all members were mailed two reminder notices asking them to complete the survey. Second, we contacted officers from all 50 state-level school psychology associations during June and July, 2017. We received approval from 34 state associations. Officers in these associations facilitated recruitment via electronic mailing lists, posts on association websites, and publication in state newsletters during July through October, 2017. Finally, participants were recruited using social media during September and October, 2017. A brief description and link to the survey were posted to six Facebook groups that specifically targeted school psychologists (viz., Said No School Psychologist Ever, School Psych Lounge, School Psych to School Psych, Sincerely School Psychologist, Notes from the School Psychologist Blog, and Get School Psyched Up!). Members of these groups were also asked to share the link with other networks of school psychologists. Posts were made in a staggered fashion, starting with the largest groups. Posts were made on Thursday at the end of the workday, as this time is suggested to be a time when users are highly engaged with Facebook (Kolowich, 2018). The largest groups were contacted a second time to recruit members from underrepresented states.

2.3. Survey design

Survey items were designed to obtain demographic information as well as information about: (a) educational and professional experiences; (b) students served (e.g., racial/ethnic groups, language proficiency, and disability status); and (c) test use and assessment procedures. Items addressing assessment methods were developed from reviews of prior assessment surveys in school psychology, recently published textbooks (e.g., Briesch, Volpe, & Floyd, 2018; Flanagan & Harrison, 2012; Kranzler & Floyd, 2013; Whitcomb, 2017), review articles (e.g., Floyd et al., 2015; Visser, Ruiter, van der Meulen, Ruijsenaars, & Timmerman, 2012), and online tool charts (e.g., National Center on Intensive Intervention, 2018). Fifteen academicians with extensive expertise in assessment in school psychology as well as other salient topics (e.g., academic and behavioral screenings, progress monitoring, diagnostic assessment, special education eligibility determination, service provision for ethnic minority students, and assessment practices and accommodations for students with disabilities) reviewed and critiqued all survey items. Items were added, deleted, and revised based on their feedback. The survey was designed, and evidence of validity accumulated, following recommendations from Johnson and Morgan (2016).

2.4. Survey administration

Qualtrics was used as the platform for delivering the survey. The survey was completed online and could be accessed using computers, tablets, or smartphones. Respondents could save their responses and return to complete the survey at their convenience. We included an item asking respondents if they had provided direct assessment, intervention, or consultation services in the past year. Anyone who responded “No” to this item was exited from the survey. Unique IP addresses were required to prevent multiple entries. The survey consisted of 65 items. Several item formats were used, including 25 multiple-choice items, 19 matrix tables, 14 sliders (where respondents drag a bar to indicate their response level), 5 text entry items, and 2 constant sum items (where numeric entries are summed). We used adaptive questioning (skip logic) to minimize response time. Participants who reported that they had not engaged in academic screenings or benchmarking, behavioral screenings, or progress monitoring skipped questions pertaining to these topics. The response period extended from May 18th to November 1st of 2017. The items, presented in online format, are available at Open Science Framework, <https://osf.io/njw69/>. The survey was open access.

2.5. Data analysis

Following the end of data collection, the full set of responses was downloaded from Qualtrics. The downloaded responses contained 1359 records, but 7 records were deleted because they were false entries from when the survey was being previewed during its development phase. A total of 1352 responses were therefore collected, including 488 fully completed questionnaires and 864 questionnaires with varying degrees of completion. All incomplete questionnaires were started, but respondents stopped responding to items at some point and exited the questionnaire.

Next, we examined the frequency of respondents who reported that they had not provided: (a) direct assessment, intervention, or consultation services as a school psychologist during the past year; or (b) direct assessment, intervention, or consultation services in a school setting for any purpose during the past year. All the respondents reported that they had provided services during the past year, although 13 reported they had not provided services in a school setting for any purpose during the past year. These 13 respondents were excluded from subsequent analyses, including examination of potential systematic bias due to missingness. Additionally, we excluded 22 individuals who did not report whether or not they had provided services in a school setting for any purpose during the past year. This process resulted in an analytic sample of 1317 responses.

The missing data pattern was *monotone* (Little & Rubin, 2002; Schafer & Graham, 2002) given that missing responses tended to be observed toward the end of the survey. Diggle and Kenward (1994) described this pattern as *completely random dropout*, which is noted as a special case of the missing completely at random (MCAR) assumption (Little, 1995; Rubin, 1976). Completely random dropout occurs when the dropout mechanism is independent of any outcome variables (i.e., observed or unobserved). Survey attrition is considered MCAR when the incomplete responses are independent of any responses given prior to the participant(s) stopping the survey. Evidence of instances when MCAR may not be tenable are when “differences in distributions of observed variables across missing data patterns” (Little & Rubin, 2002, p. 18). We collected multiple pieces of information to evaluate whether the MCAR assumption was tenable (Little, 1995).

First, we examined the distribution of an array of demographic and practice-based responses across missing data patterns using Cramer's V as a measure of effect size. Next, we computed the standardized mean difference (Cohen's d) in each of the quantitative variables across missing data patterns. We did not conduct significance tests of each of these estimates due to the large sample size. Table 4 displays the effect size estimates for relations between questionnaire completion status and background or experience variables. As can be seen here, results indicated that Cramer's $V \leq 0.1$ and Cohen's d estimates $\leq |0.2|$. Given that small effect sizes such as these are an indication that there is little to no relation between questionnaire completion status and any other reported background variables, we considered these acceptable. Thus, these analyses indicate that the absence of some responses from those participants who returned incomplete surveys did not introduce systematic bias in the results.

We then examined the relations between questionnaire completion status and the following variables: (1) gender identity; (2) age; (3) race; (4) ethnicity; (5) employment setting; (6) highest degree earned; (7) accreditation status of the degree program associated with their highest degree earned; (8) whether or not participants utilized a traditional intelligence-achievement discrepancy approach to identify specific learning disabilities; (9) whether or not they utilize a patterns of strengths and weaknesses approach to

Table 4

Effect size estimates for the relation between questionnaire completion status and background or experience variables.

Variable	Effect size estimate	
	Complete vs. incomplete	Degree of completion
	Cramer's V	Cramer's V
Gender identity	0.04	0.07
Race	0.10	0.10
Ethnicity	0.05	0.05
Setting	0.08	0.09
Highest degree earned	0.05	0.05
Graduate program approval or accreditation status	0.07	0.06
Utilize a traditional intelligence-achievement discrepancy approach to identify specific learning disabilities	0.05	0.06
Utilize a pattern of strengths and weaknesses approach to identify learning disabilities	0.08	0.06
Utilize a response to intervention approach to identify learning disabilities	0.06	0.06
	Cohen's <i>d</i>	η^2
Age	0.18	0.02
Years of experience	0.15	0.02
Approximate percentage of employment involving direct assessment, intervention, or consultation services	0.02	0.00
Approximate percentage of individuals assessed who are considered of minority status	0.06	0.02

identify learning disabilities; (10) whether or not they utilize a response to intervention (RtI) approach to identify learning disabilities; (11) years of experience; (12) the approximate percentage of employment involves direct assessment, intervention, or consultation services; and (13) the approximate percentage of students assessed who are considered of minority status. The effect size estimates are presented in Table 4. As can be seen here, all 13 relations were small enough to conclude that, in all likelihood, there were no systematic differences between the responses of those who fully completed the questionnaire and those who did not.

Due to the blocks in which questions were presented, there were incremental reductions in questionnaire completeness. Specifically, there were: 484 responses that were 100% complete, 278 that were 53–58% complete, 177 that were 40–49% complete, 220 that were 28% complete, and 158 that were 15% complete. Table 4 also displays results of the same analyses conducted on the basis of degree of completion. For the categorical variables, we again reported Cramer's *V* as the measure of effect size. For the quantitative variables, we reported η^2 as the measure of effect size. Values < 0.02 were considered acceptable. Given these results, which indicated that there were no systematic differences between the responses of those who completed the questionnaire versus those who did not, we reported all responses to each item in aggregate.

3. Results

3.1. General description of service delivery

Participants reported spending an average of about 76% (*SD* = 19%) of their time providing direct assessment, intervention, or consultation services. About 58% (*SD* = 19%) of the average school psychologist's time was spent providing services that support special education students, compared to about 16% (*SD* = 13%) of time spent providing preventive services in general education settings. Almost 29% (*SD* = 16%) of the average school psychologist's work week was spent writing and formatting reports from individualized assessments.

Elementary-age students were the most frequently served age group. On average, school psychologists appeared to spend little time working with infant, preschool, and post-secondary populations. Most participants (41%) reported suburban school districts as their primary employment setting, followed by rural (26%), urban (24%), and districts with a mix of population densities (9%). Participants indicated that about 46% of students they served were White. The percentage of students served who were Black or African American students was 28%, Biracial 9.5%, Multiracial 8%, Asian 3.5%, American Indian or Alaskan Native 2.5%, Middle Eastern or North African 2%, and Native Hawaiian or Pacific Islander < 1%. School psychologists reported conducting comprehensive assessments for suspected SLD about 25% of the time. Referral concerns for other disability categories were as follows: 18% for ADHD, 10% for ASD, 9% for language disorder, 9% for emotional disturbance, 9% for developmental delay, 7% for behavior and conduct problems, 6% for mild intellectual disability, 3% for low incidence disabilities, 3% for gifted and talented students, and 1% for students with traumatic brain injuries. Approximately 72% of participants reported that they work with students who were English Language Learners (ELL). School psychologists who reported working with this population indicated that they provided services to an average of about 2 ELL students per month (*SD* = 3.3).

3.2. Use of assessment instruments

Minimum, maximum, and average usage values, as well as the percentage of participants who used each test and procedure, are

Table 5
Average usage per month and percentage of participants using targeted assessment instruments.

Assessment instrument and overall rank by average usage per month	Type and rank	Minimum-maximum	M	SD	% Using
1 Teacher Rating Scales, BASC-3	RS 1	0–60	3.65	4.92	90.9
2 Wechsler Intelligence Scale for Children-Fifth Edition	COG 1	0–50	3.49	4.84	79.7
3 Parent Rating Scales, BASC-3	RS 2	0–40	3.10	3.62	89.9
4 Developmental history interview completed with parent or caregiver	IN 1	0–30	2.91	3.96	69.3
5 Curriculum-Based Measurement (CBM) oral reading fluency probes	ACH 1	0–77	2.88	8.55	29.3
6 CBM reading comprehension probes	ACH 2	0–51	2.22	6.10	29.5
7 Unstructured interview with child or adolescent	IN 2	0–30	2.19	3.63	50.5
8 Kaufman Test of Educational Achievement, Third Edition	ACH 3	0–60	1.94	5.05	62.0
9 CBM early literacy probes	ACH 4	0–76	1.92	6.44	26.6
10 Wechsler Individual Achievement Test, Third Edition	ACH 5	0–70	1.82	5.30	47.1
11 Self-Report of Personality, BASC-3	SELF 1	0–45	1.81	3.77	76.1
12 Woodcock-Johnson IV Tests of Achievement	ACH 6	0–55	1.78	3.95	49.7
13 CBM Math computation probes	ACH 7	0–51	1.78	5.36	27.7
14 Observation, interval recording	OB 1	0–20	1.59	2.75	60.3
15 Problem-solving interview (completed with teacher)	IN 3	0–40	1.58	3.63	52.7
16 Teacher Rating Scales, Conners 3	RS 3	0–60	1.54	3.52	42.2
17 CBM Math concept and application probes	ACH 8	0–51	1.52	4.49	27.3
18 Adaptive Behavior Assessment System, Third Edition (combined across form)	ADAPT 1	0–30	1.49	2.85	69.4
19 Parent Rating Scales, Conners 3	RS 4	0–30	1.37	2.46	61.2
20 CBM early numeracy probes	ACH 9	0–51	1.32	5.05	22.8
21 Problem-solving interview completed with parent or caregiver	IN 4	0–50	1.09	3.16	42.5
22 Observation, frequency recording	OB 2	0–20	1.07	2.09	50.4
23 Observation, A-B-C recording form	OB 3	0–30	1.03	2.21	61.7
24 Teacher Form, BRIEF2	RS 5	0–20	1.02	2.10	49.8
25 Functional assessment interview completed with teacher	IN 5	0–30	0.96	2.30	53.2
26 Beery-Buktenica Development Test of Visual-Motor Integration, Sixth Edition	NEURO 1	0–15	0.90	2.34	33.3
27 CBM written expression probes	ACH 10	0–75	0.89	4.67	16.3
28 Parent Form, BRIEF2	RS 6	0–11	0.89	1.73	48.4
29 Differential Ability Scales, Second Edition	COG 2	0–20	0.88	2.33	34.6
30 Comprehensive Test of Phonological Processing, Second Edition	ACH 11	0–50	0.85	3.01	36.9
31 Woodcock-Johnson IV Tests of Cognitive Ability	COG 3	0–20	0.85	2.00	39.3
32 Observation, duration recording	OB 4	0–20	0.74	1.94	39.9
33 Kaufman Assessment Battery for Children, Second Edition	COG 4	0–30	0.74	2.21	30.3
34 Functional assessment interview completed with parent or caregiver	IN 6	0–20	0.70	1.81	46.0
35 Behavioral Observation of Students in Schools	OB 5	0–15	0.64	1.96	21.2
36 Self-Report Form, BRIEF2	SELF 2	0–30	0.63	2.17	32.0
37 Daily Behavior Report Cards	RS 7	0–20	0.59	1.91	26.8
38 Childhood Autism Rating Scale, Second Edition	ASD 1	0–60	0.56	3.14	39.1
39 Reynolds Intellectual Assessment Scale, Second Edition	COG 5	0–35	0.56	2.34	18.3
40 Teacher Form, Vineland Adaptive Behavior Scales, Third Edition (Vineland-3)	ADAPT 2	0–10	0.54	1.15	42.6
41 Functional Assessment Observation Form	OB 6	0–35	0.52	1.98	27.9
42 Parent/Caregiver Form, Vineland-3	ADAPT 3	0–10	0.49	1.23	39.5
43 Autism Spectrum Rating Scales (combined across form)	ASD 2	0–15	0.49	1.31	31.8
44 Student Observation System, BASC-3	OB 7	0–20	0.49	1.56	22.6
45 Brief analysis of academic skills (e.g., Can't Do/Won't Do Assessment)	ACH 12	0–30	0.46	2.02	24.7
46 Portfolio, performance, or authentic assessment techniques	ACH 13	0–10	0.44	1.38	18.9
47 Gilliam Autism Rating Scale, Third Edition	ASD 3	0–10	0.43	1.09	39.7
48 Autism Diagnostic Observation Schedule, Second Edition	ASD 4	0–25	0.43	1.50	42.6
49 Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition	COG 6	0–15	0.42	1.16	35.5
50 CBM spelling probes	ACH 14	0–23	0.41	1.90	9.8
51 Teacher Form, Conners Comprehensive Behavior Rating Scales	RS 8	0–30	0.40	1.79	21.6
52 Sentence completion tests	PT 1	0–15	0.39	1.36	25.8
53 Children's Depression Inventory	SELF 3	0–20	0.38	1.41	29.4
54 Parent Form, Conners Comprehensive Behavior Rating Scales	RS 9	0–20	0.38	1.50	21.4
55 WISC-IV Integrated	NEURO 2	0–65	0.38	3.20	11.3
56 Wechsler Adult Intelligence Scale, Fourth Edition	COG 7	0–12	0.33	0.96	26.7
57 Bender Visual-Motor Gestalt Test, Second Edition	NEURO 3	0–25	0.33	1.79	10.8
58 Gray Oral Reading Tests-Fifth Edition	ACH 15	0–15	0.32	1.33	21.2
59 Wide Range Assessment of Memory and Learning, Second Edition	NEURO 4	0–10	0.31	1.03	20.2
60 Kaufman Brief Intelligence Test, Second Edition	COG 8	0–30	0.31	1.67	19.6
61 Teacher Informant, Vanderbilt Assessment Scales, Second Edition	RS 10	0–18	0.31	1.52	11.9
62 NEPSY-II	NEURO 5	0–10	0.28	1.01	21.7
63 Bracken Basic Concept Scale, Third Edition: Receptive	EARLY 1	0–20	0.28	1.51	16.9
64 Revised Children's Manifest Anxiety Scale, Second Edition	SELF 4	0–12	0.27	0.89	15.5
65 Interview Form, Vineland-3	ADAPT 4	0–10	0.26	0.84	23.3
66 Parent Informant, Vanderbilt Assessment Scales, Second Edition	RS 11	0–16	0.26	1.41	10.5
67 Conners-Wells Adolescent Self-Report Scale	SELF 5	0–15	0.26	1.03	17.1

(continued on next page)

Table 5 (continued)

Assessment instrument and overall rank by average usage per month	Type and rank	Minimum-maximum	M	SD	% Using	
68	Scholastic Reading Inventory	ACH 16	0–30	0.25	2.08	4.1
69	Wechsler Abbreviated Scale of Intelligence, Second Edition	COG 9	0–15	0.24	1.03	14.2
70	Developmental Assessment of Young Children, Second Edition	EARLY 2	0–15	0.24	1.34	22.7
71	Cognitive Assessment System, Second Edition	COG 10	0–15	0.23	1.28	8.5
72	Multidimensional Anxiety Scale for Children, Second Edition	SELF 6	0–15	0.23	1.07	15.3
73	Gifted Rating Scales	RS 12	0–35	0.22	1.83	6.7
74	House-Tree-Person	PT 2	0–10	0.22	0.89	16.3
75	Comprehensive Test of Nonverbal Intelligence, Second Edition	COG 11	0–15	0.22	0.87	26.3
76	Direct Behavior Ratings as a rating scale	RS 13	0–15	0.21	1.06	8.0
77	Observation, conditional probability recording form	OB 8	0–30	0.21	1.57	9.6
78	Direct Behavior Ratings as an observation method	OB 9	0–10	0.21	0.96	8.5
79	Teacher Form, Comprehensive Executive Function Inventory (CEFI)	RS 14	0–6	0.21	0.78	13.5
80	Stanford-Binet Intelligence Scales-Fifth Edition	COG 12	0–11	0.21	0.78	21.7
81	Parent Form, CEFI	RS 15	0–5	0.19	0.70	13.0
82	Wechsler Nonverbal Scale of Ability	COG 13	0–10	0.19	0.72	23.6
83	KeyMath-3 Diagnostic Achievement	ACH 17	0–6	0.18	0.64	18.5
84	Teacher Report Form, Achenbach System of Empirically Based Assessment (ASEBA)	RS 16	0–9	0.17	0.97	8.0
85	Draw A Person	PT 3	0–8	0.17	0.78	16.4
86	Beck Youth Inventories, Second Edition	SELF 7	0–8	0.17	0.73	11.5
87	Kinetic Family Drawing	PT 4	0–8	0.17	0.75	14.9
88	Social Responsiveness Scale, Second Edition	ASD 5	0–15	0.16	1.03	7.1
89	Battelle Developmental Inventory, Second Edition	EARLY 3	0–15	0.16	0.92	10.8
90	Child Behavior Checklist, ASEBA	RS 17	0–10	0.15	0.79	9.1
91	School Version, ADHD Rating Scale 5	RS 18	0–20	0.14	1.06	6.9
92	Teacher Form, Social Skills Improvement System-Performance Screening Guide	RS 19	0–5	0.14	0.52	13.8
93	Diagnostic interview completed with child or adolescent	IN 7	0–15	0.14	1.13	3.0
94	Bateria III Woodcock	EARLY 4	0–10	0.13	0.70	9.9
95	Scholastic Math Inventory	ACH 18	0–30	0.13	1.50	2.1
96	Parent Form, Social Skills Improvement System-Performance Screening Guide	RS 20	0–5	0.13	0.51	13.0
97	Home Version, ADHD Rating Scale 5	RS 21	0–20	0.12	1.03	5.1
98	Reynolds Adolescent Depression Scale, Second Edition	SELF 8	0–15	0.12	0.80	9.8
99	Delis-Kaplan Executive Function System	NEURO 6	0–15	0.11	0.81	9.2
100	Peabody Picture Vocabulary Test, Fourth Edition	ACH 19	0–8	0.11	0.71	5.7
101	Universal Nonverbal Intelligence Test, Second Edition	COG 14	0–5	0.10	0.42	17.9
102	Developmental Indicators for the Assessment of Learning, Fourth Edition	EARLY 5	0–9	0.10	0.78	2.8
103	Beck Depression Inventory, Second Edition	SELF 9	0–5	0.10	0.41	12.5
104	Bayley Scales of Infant and Toddler Development	EARLY 6	0–7	0.10	0.65	5.9
105	Diagnostic interview completed with parent or caregiver	IN 8	0–15	0.09	0.94	3.4
106	Test of Memory and Learning, Second Edition	NEURO 7	0–8	0.08	0.56	6.0
107	Gilliam Asperger's Disorder Scale	ASD 6	05	0.08	0.37	9.7
108	Youth Self-Report, ASEBA	SELF 10	0–8	0.07	0.55	5.1
109	Iowa Test of Basic Skills	ACH 20	0–6	0.07	0.47	4.5
110	Informal Reading Inventory	ACH 21	0–5	0.07	0.47	5.3
111	Roberts Apperception Test for Children, Second Edition	PT 5	0–10	0.07	0.51	7.9
112	Reynolds Child Depression Scale, Second Edition	SELF 11	0–5	0.07	0.35	7.1
113	BRIGANCE Inventory of Early Development II	EARLY 7	0–8	0.06	0.46	5.1
114	Woodcock Reading Mastery Tests, Third Edition	ACH 22	0–6	0.06	0.38	4.3
115	Beck Anxiety Inventory	SELF 12	0–10	0.06	0.50	5.1
116	Suicidal Ideation Questionnaire	SELF 13	0–6	0.06	0.41	5.3
117	Self-Report, Social Skills Improvement Rating Scales	SELF 14	0–5	0.06	0.39	7.1
118	Thematic Apperception Test	PT 6	0–10	0.06	0.53	5.1
119	Continuous performance tests	NEURO 8	0–5	0.05	0.37	3.7
120	Motor-Free Visual Perception Test-4	NEURO 9	0–5	0.05	0.36	4.3
121	Social, Academic, and Emotional Behavior Risk Screeners	RS 22	0–15	0.05	0.73	1.7
122	Teacher Rating Scales, Clinical Assessment of Behavior	RS 23	0–10	0.05	0.57	1.3
123	Leiter International Performance Scale, Third Edition	COG 15	0–5	0.05	0.28	7.9
124	Stroop Color and Word Test	NEURO 10	0–10	0.04	0.53	2.1
125	Parent Rating Scales, Clinical Assessment of Behavior	RS 24	0–10	0.04	0.52	1.1
126	Attention Deficit Hyperactivity Disorder School Observation Code	OB 10	0–15	0.04	0.69	1.1
127	Dean-Woodcock Sensory-Motor Battery	NEURO 11	0–15	0.03	0.69	0.9
128	Millon Adolescent Clinical Inventory	SELF 15	0–5	0.03	0.28	2.8
129	School Social Behavior Scales	RS 25	0–5	0.03	0.31	2.9
130	Rorschach Inkblot Test	PT 7	0–10	0.03	0.46	1.7
131	Scales of Independent Behavior-Revised	ADAPT 5	0–5	0.03	0.25	3.2
132	Personality Inventor for Children, Second Edition	SELF 16	0–5	0.03	0.25	2.8
133	Direct Observation Form, ASEBA	OB 11	0–5	0.02	0.26	2.5
134	Reynolds Intellectual Screening Test	COG 16	0–5	0.02	0.27	1.5
135	Devereux Student Strengths Assessment	RS 26	0–5	0.02	0.26	1.3
136	Minnesota Multiphasic Personality Inventory-Adolescent, Restructured Form	SELF 17	0–5	0.02	0.25	1.5

(continued on next page)

Table 5 (continued)

Assessment instrument and overall rank by average usage per month		Type and rank	Minimum-maximum	<i>M</i>	<i>SD</i>	% Using
137	Wisconsin Card Sorting Test	NEURO 12	0–5	0.02	0.25	5.6
138	State-Event Classroom Observation System	OB 12	0–5	0.01	0.24	0.9
139	State-Trait Anxiety Inventory for Children	SELF 18	0–5	0.01	0.24	0.9
140	Kaufman Adolescent and Adult Intelligence Test	COG 17	0–5	0.01	0.23	0.8
141	Halstead-Reitan Neuropsychological Test Battery	NEURO 13	0–5	0.01	0.23	0.4
142	Classroom Assessment Scoring System	OB 13	0–5	0.01	0.23	0.4
143	Systematic Screening of Behavior Disorders, Second Edition	OB 14	0–3	0.01	0.15	1.1
144	Revised Edition of the School Observation Coding System	OB 15	0–3	0.01	0.17	0.6
145	Young Children's Achievement Test	ACH 23	0–3	0.01	0.14	0.4
146	Impairment Rating Scale	RS 27	0–3	0.01	0.14	0.4
147	Mullen Scales of Early Learning	EARLY 8	0–0.25	0.00	0.02	0.6
148	Boehm Test of Basic Concepts, Third Edition	EARLY 9	0–0.50	0.00	0.03	0.4
149	Learning Propensity Assessment Device	EARLY 10	0–0.25	0.00	0.01	0.2
150	Merrill-Palmer, Revised Scales of Development	EARLY 11	0–0.25	0.00	0.01	0.2

Note. RS = rating scale, COG = cognitive ability test, IN = interview, ACH = academic achievement test, SELF = self-report scale, OB = observation, ADAPT = adaptive behavior scale, ASD = autism spectrum disorder scale, NEURO = neuropsychological test, PT = projective test or technique, EARLY = early childhood/developmental assessment.

presented in Table 5. This table includes all instruments listed in the survey as well as write-in responses reported at least five times in text boxes. Although we report the most recent editions of tests and other assessment tools, responses are inclusive of previous editions in cases where participants may not have adopted the most recent version. Instruments are ranked by average usage per month and by instrument type in this table. As this table shows, Teacher Rating Scales (TRS) and Parent Rating Scales (PRS) of the BASC-3 (Reynolds & Kamphaus, 2015a) were two of the three most frequently used instruments. The Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V; Wechsler, 2014) was the second most used instrument. Developmental history interviews ranked fourth. Previous surveys did not ask about these interviews, but as our results indicate, school psychologists routinely conduct them. The fifth and sixth most frequently used instruments were standardized general achievement outcome measures, commonly referred to as Curriculum-Based Measurement (CBM), which assess students' oral reading fluency (R-CBM) and reading comprehension (MAZE), respectively. Unstructured interviews also ranked highly (seventh). In contrast to previous surveys that considered student, teacher, and parent interviews as a whole, this ranking only reflects interviews with students. Also represented in the Top 10 are early literacy probes and two standardized, broad-band tests of academic achievement. Results within each test and procedure category are discussed next by category.

3.3. Behavior rating scales

In stark contrast to previous surveys, 97% of participants reported that they use behavior rating scales. As shown in Table 5, 90% of participants reported using the BASC-3 rating scales. The BASC-3 TRS were administered an average of 3.65 ($SD = 4.92$) times per month, while BASC-3 PRS were administered 3.10 ($SD = 3.62$) times per month. Given that 18% of referrals relate to concerns with ADHD symptoms, it is not surprising that the Conners 3 Teacher and Parent Scales (Conners, 2008) were each administered an average of about 1.5 times per month ($SD = 3.52$ and $SD = 5.05$, respectively). Likewise, the Behavior Rating Inventory of Executive Function, Second Edition (BRIEF2) Teacher and Parent Forms (Gioia, Isquith, Guy, & Kenworthy, 2015) were used regularly ($M = 1.02$, $SD = 2.10$ and $M = 0.89$, $SD = 1.73$). Last, daily behavior report cards (Volpe & Fabiano, 2013) were administered relatively frequently ($M = 0.59$, $SD = 1.91$).

In addition to these results, about 20% of participants indicated that they administer or interpret behavioral rating scales in a universal screening process for behavioral concerns. Most participants who screen (89%) reported that they were involved in three behavioral screening applications annually, with an average of 350 students per application. The Behavioral and Emotional Screening System (BASC-3 BESS; Kamphaus & Reynolds, 2015) was reported to be the most widely used behavioral screening measure. About 68% of those who engaged in behavioral screening reported using the BASC-3 BESS, with 35% using it in isolation, and 23% using it in combination with one or two additional screening instruments. About 25% of those who engage in screening reported that they use the Student Risk Screening Scale (Drummond, 1994) with about 18% indicating they use it in isolation. About 11% of participants reported using the Social, Academic, and Emotional Behavior Risk Screener (Kilgus, Eklund, von der Embse, Taylor, & Sims, 2016), with 9% indicating they use it as a standalone measure.

3.3.1. Adaptive behavior scales¹

About 88% of participants reported using adaptive behavior scales, most of which utilize a rating scale format. The Adaptive

¹ Although most adaptive behavior rating scales use a rating scale format, some utilize an interview format. We distinguish these measures from behavior rating scales based on conceptual differences and tradition. Specifically, the measures we labeled as adaptive behavior scales are those that specifically target “the level of everyday performance of tasks that is required for a person to fulfill typical roles in society, including maintaining independence and meeting cultural expectations regarding person and social responsibility” (VandenBos, 2006, p. 18).

Behavior Assessment System, Third Edition (ABAS-3; Harrison & Oakland, 2015) was the most widely used instrument ($M = 1.49$, $SD = 2.85$) in this category. The ABAS-3 thus surpasses the most widely used adaptive behavior scale in prior surveys, the Vineland Adaptive Behavior Scales, Third Edition (Vineland-3; Sparrow, Cicchetti, & Saulnier, 2016). The Vineland Teacher, Parent, and Interview forms were used an average of about once per month. Adaptive behavior rating scales other than the ABAS-3 and Vineland-3 were rarely administered.

3.3.2. Self-report forms

Further underscoring its popularity, the most frequently used self-report measure was the Self-Report of Personality (SRP; $M = 1.81$, $SD = 3.77$) of the BASC-3. Approximately 84% of participants reported using self-report forms, with most relying primarily on the SRP. Approximately 76% of participants reported that they use the SRP, which is a broad-band measure of personality (Whitcomb, 2017). Other broad-band self-report forms were used far less frequently. Only 17% and 5% of participants reported using the next two most popular broad-band self-report forms, the Conners-Wells Adolescent Self-Report Scale ($M = 0.26$, $SD = 1.03$; Conners, 1997) and Beck Youth Inventories, Second Edition ($M = 0.17$, $SD = 0.73$; Beck, Beck, Jolly, & Steer, 2005), respectively. The second most frequently used self-report form was the BRIEF2 Self-Report Form ($M = 0.63$, $SD = 2.17$; Gioia et al., 2015). The fact that 32% of participants used this measure indicates an increase in the assessment of executive functioning. Participants reported using several narrow-band measures of depression and anxiety, although all were used sparingly.

3.3.3. Behavioral observations

As is also shown in Table 5, interval recording was the most commonly used observational method ($M = 1.59$, $SD = 2.75$), followed by frequency recording ($M = 1.07$, $SD = 2.09$). Duration recording also was used relatively frequently ($M = 0.74$, $SD = 1.94$). Participants reported using antecedent-behavior-consequent (ABC) recording forms an average of about once per month ($SD = 2.21$). The fifth most popular observational method was a structured observation tool, the Behavioral Observation of Students in Schools (BOSS; Shapiro, 2004). The BOSS was used about 0.64 times per month ($SD = 1.96$). Although several other structured observation forms were reported to be used by participants, most were employed sparingly.

3.3.4. Interviews

Developmental history interviews ($M = 2.91$, $SD = 3.96$) and unstructured interviews with students ($M = 2.19$, $SD = 3.63$) were the most frequently used interviews. Participants also reported frequent use of problem-solving interviews with teachers ($M = 1.58$, $SD = 3.63$) and parents/caregivers ($M = 1.09$, $SD = 3.16$). In addition, school psychologists indicated using functional behavioral assessment interviews often, but diagnostic interviews were used sparingly by all but a handful of participants.

3.3.5. Tests of academic achievement

Approximately 85% of school psychologists reported that they administered at least one test of academic achievement. In contrast to the results of prior surveys, R-CBM probes ($M = 2.88$, $SD = 8.55$) and MAZE probes ($M = 2.22$, $SD = 6.10$) were the two most frequently used measures of academic achievement. In addition, early literacy probes, mathematics computation probes, mathematics concept and application probes, and early numeracy probes were among the 20 most used assessment measures. Although written expression probes were not used as often as reading and mathematics probes, they were administered an average of about once per month ($M = 0.89$, $SD = 4.67$). The percent of participants who reported using the various CBM probes ranged from 16% for written expression to just under 30% for R-CBM and MAZE.

About 46% of participants indicated that they use academic assessment instruments for progress monitoring purposes, and about 49% indicated that they administered or interpreted instruments in a universal screening/benchmarking process for academic concerns. Most of these participants (74%) reported involvement in three screening applications annually, with a median of 450 students in each application. The number of students screened varied based on the size of the district served. It is difficult to estimate usage precisely, however, as probes could be administered, scored, or interpreted by school personnel other than school psychologists.

About 18% of participants reported the names of vendors offering CBM probes for academic screening and progress monitoring. Most of this group (53%) reported drawing probes from multiple vendors; AIMSweb and DIBELS probes were used in combination by 12%. Approximately 48% reported administering AIMSweb probes, with 21% using them in isolation and 27% using them in combination with probes from other vendors. About 36% reported administering DIBELS probes, with only 8% using them in isolation and 28% using them in combination. The next most commonly used vendors included easyCBM (23%), Renaissance Star (14%), NWEA (11%), FastBridge (9%), and iReady (9%).

Following use of CBM, the third most frequently administered test of academic achievement was the Kaufman Test of Educational Achievement, Third Edition (KTEA-3; Kaufman & Kaufman, 2014), followed closely by the Wechsler Individual Achievement Test, Third Edition (WIAT-III; Wechsler, 2009) and the WJ IV Tests of Achievement (Schrank, McGrew, & Mather, 2014). These three tests were ranked as the 8th, 10th, and 12th most frequently used assessment instruments. On average, school psychologists administered about two of each per month, although there was a great deal of variability in reported frequency.

In addition to CBM probes and broad-band, norm-referenced tests of academic achievement, school psychologists also reported using other specialized measures of academic skills. On average, respondents administered about one Comprehensive Test of Phonological Processing, Second Edition (Wagner, Torgesen, Rashotte, & Pearson, 2013) per month ($M = 0.85$, $SD = 3.01$). Brief experimental analysis of academic skills (e.g., Can't Do/Won't Do Assessment; Daly, Witt, Martens, & Dool, 1997) was used relatively infrequently ($M = 0.46$, $SD = 2.02$), as were portfolio, performance, or authentic assessment techniques ($M = 0.44$, $SD = 1.38$). On

average, school psychologists used specialized measures such as the Gray Oral Reading Tests, Fifth Edition (Wiederholt & Bryant, 2012), the Scholastic Reading Inventory (Houghton Mifflin Harcourt, 2011), and the KeyMath-3 Diagnostic Achievement Test (Connolly, 2007) approximately two to three times per year.

3.3.6. Tests of cognitive ability

Over 95% of participants reported that they administered at least one test of cognitive ability during the past year. According to our results, the WISC-V was used more frequently than the next five most used cognitive tests combined. The WISC-V was administered 3.49 times per month ($SD = 4.84$), on average. The next most frequently administered test of cognitive ability, the Differential Ability Scales, Second Edition (DAS-2; Elliot, 2007), was given an average of 0.88 times per month ($SD = 2.33$). The next three most frequently administered tests of cognitive ability were the WJ IV Tests of Cognitive Ability (Schrank et al., 2014), the Kaufman Assessment Battery for Children, Second Edition (Kaufman & Kaufman, 2004), and the Reynolds Intellectual Assessment Scale, Second Edition (Reynolds & Kamphaus, 2015b). These instruments were administered an average of 0.85 ($SD = 2.00$), 0.74 ($SD = 2.21$), and 0.56 ($SD = 2.34$) times per month, respectively.

3.3.7. Projective instruments

Approximately 32% of participants reported administering at least one projective instrument during the past year. Sentence completion tests were the most widely used projective measures, reportedly used by 26% of school psychologists. About 93% of those who used these tests did so once per month or less ($M = 0.39$, $SD = 1.36$), however. The House-Tree-Person was administered by about 17% of school psychologists, with 97% of those reporting indicating that they used it once per month or less. Other projective drawing tests, such as the Draw-A-Person test and the Kinetic Family Drawing were given slightly less frequently than the House-Tree-Person. Similar to the results of prior survey research, apperception tests and the Rorschach Inkblot Test were rarely used. Thus, although projective tests tend to be used less frequently than in the past, when they were used, they continue to consist of quick and easy-to-use projective tests rather than more time-consuming ones.

3.3.8. Tests of early development

Our results indicate that most school psychologists (62%) do not administer tests of early development. For those who did, the most frequently used measure is the Bracken Basic Concepts Scale, Third Edition: Receptive (BBCS-3:R; Bracken, 2006), which about 17% of participants reported using. About 95% of those who used it did so one or less times per month, however. The Developmental Assessment of Young Children, Second Edition (Voress & Maddox, 2012) was used just slightly less than the BBCS-3:R. About 11%, 6%, and 3% of participants reported using the Battelle Developmental Inventory, Second Edition (Newborg, 2005), the Bayley Scales of Infant and Toddler Development (Bayley, 2006), or and the Developmental Indicators for the Assessment of Learning, Fourth Edition (Mardell & Goldenberg, 2011), respectively. Other tests of early development were rarely used.

3.3.9. Neuropsychological tests

Approximately 55% of participants reported using neuropsychological tests, although our results indicate that school psychologists seldom administered traditional neuropsychological tests explicitly designed to be sensitive to brain lesions and discriminate among individuals with or without brain injuries. For example, only two participants ($< 0.01\%$) reported using the Halstead-Reitan Neuropsychological Battery (Reitan, 1979) and only four and six participants ($< 0.01\%$) reported using the Dean-Woodcock Sensory-Motor Battery (Dean & Woodcock, 2003) and the Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993), respectively. The most frequently used measure in this category was the Beery-Buktenica Development Test of Visual-Motor Integration, Sixth Edition (Beery VMI; Beery, Buktenica, & Beery, 2010), which was administered an average of once per month. School psychologists reported administering the WISC-V Integrated (Wechsler, 2015); the Bender Visual-Motor Gestalt Test, Second Edition (Brannigan & Decker, 2003); the Wide Range Assessment of Memory and Learning, Second Edition (Sheslow & Adams, 2003); and the NEPSY-II (Korkman, Kirk, & Kemp, 2007) about two to four times per year, on average. In addition, tests of continuous performance were rarely administered.

3.3.10. Instruments for assessing symptoms of ASD

About 82% of participants reported using instruments designed to assess symptoms of ASD. The Childhood Autism Rating Scale, Second Edition (Schopler, Van Bourgondien, Wellman, & Love, 2010) was the most commonly used instrument for the assessment of ASD ($M = 0.56$, $SD = 3.14$), followed closely by the Autism Spectrum Rating Scales (ASRS; Goldstein & Naglieri, 2009), which includes parent/caregiver and teacher rating scales. The ASRS was used an average of 0.49 times per month ($SD = 1.31$). The Gilliam Autism Rating Scale, Third Edition (Gilliam, 2013) and the Autism Diagnostic Observation Schedule, Second Edition (Lord et al., 2012) were both administered an average of 0.43 times per month ($SD = 1.09$ and $SD = 1.50$, respectively).

3.4. Use of assessment technology

When asked to estimate how often they used software or online services to score different types of instruments, 97% of participants reported using these methods at least occasionally. As can be seen in Table 6, about 94% reported using technology-based methods to score behavior rating scales, and 47% reported using these methods exclusively. Furthermore, more than two-thirds of participants reported using software or online services to score tests of cognitive ability and academic achievement, adaptive behavior scales, and self-report rating scales. Behavior observations, developmental tests, interviews, and neuropsychological tests were

Table 6
Online or software-based technology for scoring and administration by instrument type.

Instrument type	Scoring			Administration		
	Percent using	Percent using exclusively	Average percent use (SD)	Percent using	Percent using exclusively	Average percent use (SD)
Behavior rating scales	94.3	47.1	77.9 (31.5)	54.9	9.6	29.4 (37.5)
Self-report forms	68.7	29.8	47.9 (43.8)	40.3	6.3	16.4 (31.4)
Behavior observations	40.5	6.9	17.5 (33.1)	38.2	5.3	15.2 (31.0)
Interviews	24.3	1.4	3.6 (14.9)	24.9	1.7	4.7 (17.8)
Academic achievement tests	76.0	28.5	57.7 (43.2)	32.0	4.1	14.1 (30.5)
Cognitive ability tests	85.6	31.6	65.0 (39.1)	47.8	5.5	26.0 (38.0)
Adaptive behavior scales	74.3	48.8	59.9 (45.9)	39.3	10.7	18.7 (34.8)
Developmental tests	30.2	7.1	10.4 (28.0)	21.0	1.4	2.4 (12.7)
Neuropsychological tests	38.6	10.7	17.9 (34.8)	26.9	4.0	7.67 (23.8)

the least likely to be scored using software or online services, but approximately a quarter or more of participants reporting using these methods.

When asked to estimate how often they used software or online services to administer instruments, almost two-thirds (64%) of participants reported using these methods at least occasionally. As can also be seen in Table 6, 54.9% reported using these methods to administer behavior rating scales, but few (9.6%) reported using them exclusively. All other instrument types were administered less than half the time via software or online services, including tests of cognitive ability, self-report forms, adaptive behavior scales, and behavior observations. Fewer than one-third of participants reported administering academic achievement tests, neuropsychological tests, interviews, and developmental tests using software or online services.

4. Discussion

A number of surveys have been conducted over the years to determine the psychological and educational tests and practices employed by school psychologists when conducting assessments in the schools (Goh et al., 1981; Hutton et al., 1992; Reschly et al., 1987; Stinnett et al., 1994; Wilson & Reschly, 1996). However, the most recent of these surveys was published over 20 years ago. The purpose of the current study, therefore, was to examine the test use and assessment practices of contemporary school psychologists in the United States.

4.1. Dramatic changes

In contrast to the limited change from 1980 to the early 1990s, results of our survey clearly show that a radical shift in test usage and assessment practices has occurred over the past two decades. The most pronounced change is the increased use of broad-band behavior rating scales, particularly the BASC-3. The BASC TRS and PRS were the first and third most widely used instruments, with approximately 90% of all school psychologists reporting that they administered them. The popularity of these instruments is understandable given that they are very time- and cost-efficient tools for measuring a wide range of child's adaptive and problem behaviors in different settings and with different informants. In addition to the widespread use of broad-band rating scales, narrow-band rating scales are also widely used to assess functioning in specific areas. One example of such scales is the BRIEF2, which is used to assess executive functioning deficits associated with ADHD. Rating scales have also become the most widely used approach for the assessment of adaptive behavior. The ABAS-3 is currently the most widely used measure of adaptive behavior – although it ranks only 18th in Table 5. The rankings for measures of adaptive behavior suggest that the assessment of adaptive behavior is not part of a standard comprehensive evaluation and occurs only when ID or developmental delays are suspected.

Another major change in test usage and assessment practices is in the area of personality assessment. Two decades ago, the most widely used instruments in this area were easy-to-use projective drawings, such as the Draw-A-Person test, even though they had long been criticized on psychometric grounds (e.g., Kamphaus & Pleiss, 1991). Today, few projective tests of any kind are used. Our results are similar to those of Hojniski, Morrison, Brown, and Matthews (2006), who found that 38% of participants used projective tests, with sentence completion tests being the most frequently administered measure. We found that less than one-third of participants reported administering projective tests and that nearly all used them less than once per month. In addition, tests of perceptual and motor functioning, once a mainstay of comprehensive evaluations, are no longer routinely administered by most school psychologists. In prior surveys, the Bender-Gestalt was typically ranked as the third or fourth most widely used instrument as a measure of visual-motor functioning, personality, or both. Results of our study indicate that contemporary school psychologists rarely administer these tests for any purpose.

The assessment of academic achievement has also changed dramatically. Three of the Top 10 most used tests in our study were curriculum-based measures of reading (viz., fluency, comprehension, and early literacy). The increased use of more direct and frequent assessments of academic achievement is likely a result of the use of screening and progress monitoring in school-based MTSS. Nevertheless, similar to Stinnett et al. (1994), fewer than 30% of all school psychologists reported using CBM. This percentage is lower than that reported by Shapiro et al. (2004), who found that 54% of a 1999–2000 sample reported using CBM. The

discrepancy may be related to the inclusion of a general usage question in their survey, whereas we asked about usage of various types of CBM probes separately. In any case, either CBM has not been as widely adopted as anticipated or those measures are being administered by school personnel other than school psychologists (e.g., teachers and instructional support personnel). Further research is needed to answer this question. In addition to the use of CBM, standardized norm-referenced tests of academic achievement are also reported to be widely used by school psychologists.

Prior surveys did not inquire about school psychologists' use of computer scoring and administration technology because they were not widely available for use at that time. As the results of our study show, at the current time, there is widespread use of test scoring via software or online services. Technology use was most commonly used across norm-referenced rating scales targeting parent, teacher, and self-reports, but it was also used with norm-referenced tests of cognitive ability and achievement. Many school psychologists reported that they exclusively scored these instruments using these services. Although school psychologists reported that they administered tests less frequently than they scored them using software or online services, about two-thirds reported doing so on occasion. The most frequent use of these services was for the administration of behavior rating scales and tests of cognitive ability, but few did so exclusively. Nevertheless, as our findings show, there have been substantial changes from the days of using only physical test materials (including test manipulatives and printed copies of all forms) and scoring based on hard copies of norms tables. We welcome this application of technology as it may yield more precise measurements during testing (e.g., response times) and reduce administration and scoring errors. Although not fully realized at present, computer scoring has the potential to produce refined index scores that provide a more precise measurement of the constructs they are intended to assess than current hand-scored methods (e.g., see Benson, Kranzler, & Floyd, 2016).

4.2. Stability over time

In contrast to these dramatic changes in test usage and assessment practices, one area has remained largely unchanged – intellectual assessment. In fact, over the past 40 years, the Wechsler Intelligence Scales have consistently been one of the most widely used instruments overall by school psychologists. The Wechsler Scales are clearly the “gold standard” for measuring intelligence in school psychology. This is quite surprising given that all intelligence tests primarily measure the same thing, namely, general cognitive ability, or psychometric *g* (e.g., Floyd, Reynolds, Farmer, & Kranzler, 2013). Canivez and Watkins (2016) found that psychometric *g* accounted for over 80% of the total true score variance on the WISC-V. They also found that, of the five composite scores, only one (Processing Speed) accounted for a sufficient amount of true score variance to warrant clinical interpretation. Given the fact that the WISC-V measures the same or highly similar constructs of CHC theory that are measured by many other intelligence tests, in addition to psychometric *g*, an explanation for its long reign as the most widely used instrument in the area of intelligence testing is currently lacking.

Two other areas that have changed little over the past 20 years are neuropsychological and early childhood/developmental assessments. Large standard deviations, however, suggest considerable variability in the use of neuropsychological instruments. This means that most school psychologists do not administer neuropsychological instruments, but that those who do, do so regularly. The same pattern of results was observed with early childhood/developmental assessment. The finding that few school psychologists are engaged in testing of very young children is consistent with the results of prior survey research on test usage and administration practices (e.g., Stinnett et al., 1994).

In addition to the use of rating scales and standardized and non-standardized tests, results of our survey also indicated that school psychologists routinely conduct interviews and observations. Conducting developmental history interviews with parents or caregivers and unstructured interviews with a child or adolescent were both ranked in the Top 10 most used assessment practices. Problem-solving interviews and functional assessment interviews with teachers also were commonly administered. In addition, the majority of school psychologists reported that they regularly conducted a number of different kinds of structured observations, including systematic interval-based and frequency recording and functional behavior assessment observations. Structured behavioral observation did not rank as highly as in Wilson and Reschly's (1996) survey, largely due to the fact that we divided it into different types of observation. Taken as a whole, however, it is clear that school psychologists are engaged in a considerable number of behavioral observations, perhaps as a function of the passage of PL 105–17 in 1997, which mandated functional behavioral assessments when a student's behavior interferes with the educational process.

4.3. Strengths and limitations

Our survey's sample size was large in comparison to prior surveys of test use and assessment practices in school psychology (Goh et al., 1981; Hutton et al., 1992; Reschly et al., 1987; Stinnett et al., 1994; Wilson & Reschly, 1996). In addition, the demographic characteristics of participants in this study are comparable to those of a recent survey of NASP members (Walcott & Hyson, 2018). Nonetheless, in addition to sending the survey to 1000 randomly selected NASP members, we also recruited participants from state organizations and from six Facebook groups for school psychologists. Due to these multiple sampling methods, we were unable to determine the response rate to our survey, so the influence of nonresponse error on our results is unknown. Furthermore, it is possible that at least some of the participants in this study are members of their state organizations or of Facebook groups but not members of NASP. Thus, care should be taken when generalizing the results of our study to all NASP members. Conversely, a comparison of NASP membership data with state-level data suggest that half or fewer contemporary school psychologists are NASP members (e.g., Fagan & Wise, 2007; Jimerson, Skokut, Cardenas, Malone, & Stewart, 2008). Thus, our sampling methods likely produced results that are more representative of school psychology assessment practices across the United States than if our survey had targeted only NASP

members.

The comprehensive nature and structure of our survey also may have affected our results. Completed surveys were obtained from 488 participants, and partially completed surveys were obtained from 864 participants. We also excluded surveys from 35 individuals who indicated that they had not delivered any school psychological services (assessment, intervention, or consultation) over the past year. Prior to data aggregation and analysis, we examined whether responses from those who completed the entire survey differed systematically from those with incomplete surveys. Results of analyses of the MCAR assumption supported the conclusion of completely random dropout, we reported results of analyses for the aggregated data. However, as MCAR is generally considered a strong assumption, it is possible that some unexamined variable not included in this study may have influenced results. Future research in this area may benefit from the use of shorter and more narrowly focused surveys (e.g., contemporary intelligence test use and interpretation) and enhanced technology, such as contemporaneous reporting applications.

Because the aim of this study was to examine contemporary test use and assessment practices of school psychologists, we asked respondents to report the number of times they used specific tests and assessment practices per month. It is important to note that we did not ask them to report on how these tests and assessment practices were being used. For example, although the overall frequency with which projective drawings are used today has declined sharply from over 20 years ago, some school psychologists still report using them. We do not know, however, whether they are being interpreted according to the projective hypothesis (e.g., Lilienfeld, Wood, & Garb, 2000) or as “ice breakers” to establish rapport and facilitate conversation. With regard to intellectual assessment, results of a recent survey by Sotelo-Dynega and Dixon (2014) found notable differences in how intelligence testing was conducted by school psychologists, particularly with regard to the use of cross-battery assessment methods (e.g., see Flanagan, Ortiz, & Alfonso, 2013). Given the existence of different interpretive frameworks for assessment in many areas, such as intelligence testing, further research is needed to examine how different tests and practices are used in practice.

Last, every method of scientific inquiry is subject to limitations. Survey research is no exception. In addition to their many advantages, such as convenient data collection, low cost, and high sample representativeness, potential limitations of surveys include dishonest responding and lack of conscientiousness, among others. Although our use of anonymous responding likely reduced or eliminated social desirability bias (i.e., answering questions in a way that will be viewed favorably by others) and the desire to protect one's privacy, both of which are forms of dishonest responding, there is no way for us to know the degree to which respondents answered questions conscientiously. In our study, examination of the minimum and maximum values for the frequency of test use and assessment practices per month suggests that a very small number of participants did not carefully read the instructions to all questions before responding. For example, for the WISC-V, although descriptive statistics for its use per month seem reasonable, the minimum value is 0 and the maximum value is 50. In contrast to high maximum values for some other tests, such as the BASC TRS and PRS, which are completed by teachers and parents, respectively, school psychologists must individually administer the WISC-V. The administration of 2 or more intelligence tests per day, although certainly possible, seems unusually high and raises the question of whether some respondents reported frequency of test use per year instead of per month. Because we have no way of knowing the accuracy of responses, results of this study are based on the entire sample. Given the possibility of some spurious scores, descriptive statistics for some tests and assessment practices may be slightly inflated.

4.4. Implications for training and research

First, at this time, broad-band rating scales are the most widely used measures of adaptive and problem behaviors in the schools. Broad-band rating scales measure multiple areas of concern simultaneously, which increases the likelihood of detecting co-morbid disorders. Nonetheless, compared to externalizing problems (e.g., aggression), the covert nature of internalizing problems presents challenges to teachers and parents rating the behavioral signs of anxiety and depression. This underscores the need to supplement behavior rating scales with self-report instruments. Although the results of this study indicated that child and adolescent self-report scales are indeed administered by school psychologists, the most widely used instrument is an omnibus measure of personality, the BASC SRP. In contrast, narrow-band instruments, which focus on the assessment of single disorders, are used relatively sparingly. For example, narrow-band measures of depression and anxiety, such as the Children's Depression Inventory (CDI; Kovacs, 2010) and the Revised Children's Manifest Anxiety Scale, Second Edition (RCMAS-2; Reynolds & Richmond, 2008), are administered by school psychologists less than once per month, on average. Because self-report narrow-band instruments can be quite useful in these situations, they are likely under-utilized by school psychologists and perhaps not emphasized enough during training at this time.

Second, although school psychologists regularly conduct developmental history interviews with parents or caregivers and interviews with teachers and students, these interviews tend to be unstructured. Results of our survey revealed that very few school psychologists currently conduct structured or semi-structured diagnostic interviews. This is likely due to the fact that these interviews tend to be more time-intensive to administer and, because they are designed to identify psychopathology as defined by *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013), have limited relevance for the classification of disabilities under IDEA. Nonetheless, structured and semi-structured interviews tend to be more reliable than unstructured interviews, measure aspects of behavior not tapped by rating scales (e.g., onset and duration of problem behavior), and facilitate accurate diagnosis. Thus, given these advantages over unstructured interviews, research is needed to either adapt existing structured or semi-structured interviews – such as the Kiddie Schedule for Affective Disorders and Schizophrenia for School-Aged Children-Present-Lifetime version (K-SADS-PL; Kaufman et al., 1997) – or develop new versions of such instruments for use in educational contexts.

Third, results of our survey indicated that many school psychologists score tests via software or online services. At the current time, those who do so report that they exclusively score broad-band behavior rating scales, such as the BASC-3, using software and online services, but they also use them for scoring tests of intelligence and achievement. A majority of school psychologists also

reported that they regularly administer rating scales and tests using technology (e.g., computer or tablet). At present, however, virtually no independent, peer-reviewed research has been conducted on the comparability of instruments administered in the traditional face-to-face manner and via computer technology. Although this would appear to be less of a potential problem for behavior rating scales, because they are based on a rater's impressions of the subject's behavior, it may be of considerable concern for tests that require the establishment of rapport, querying of vague responses, scoring of items that require judgment, and administering tests that involve manipulatives. Research in this area is greatly needed before the field moves further toward computerized test administration and scoring.

Fourth, and last, further research is needed on the content and scope of assessment curricula in school psychology training programs. Because the focus of our study was on the test use and assessment practices of practitioners in the field, we did not survey faculty in school psychology training programs. To the best of our knowledge, the only comprehensive assessment survey that has been conducted on both practitioners and faculty was done by Wilson and Reschly (1996), and they found a high degree of congruity between the content of training and practitioners' assessment practices. Since then, as the results of our survey show, test use and assessment practices have changed dramatically in some areas and stayed largely the same in others. Results of our survey do not shed light on whether school psychology programs are currently meeting the training needs of future practitioners with regard to assessment. Moreover, results of our survey do not explain why the observed changes in the field occurred. Did they result from changes to training, legislation, or something else (e.g., the evidence-based practice movement)? Further research is needed to answer these questions.

5. Conclusion

The test use and assessment practices of school psychologists have evolved significantly over the past 20 years. Much of this change consists of tests and assessment practices with greater psychometric support replacing those with limited reliability and validity. As a result, today, social and emotional functioning, personality, and adaptive behavior are primarily assessed using behavior rating scales. In addition, the assessment of academic achievement is no longer conducted solely as a summative evaluation with norm-referenced standardized tests. Currently, CBM is also widely used to provide a more direct assessment of achievement and to monitor progress as a formative evaluation. In addition, intelligence testing, behavioral observation and functional behavioral assessment, and interviews continue to be staples of assessments conducted by school psychologists.

Since the authorization of IDEA in the 1970s, school psychologists have routinely served as integral members of the multi-disciplinary team that determines eligibility for special education and related services. They are largely responsible for conducting the comprehensive psychoeducational evaluations for eligibility determination. Results of this survey indicate that most school psychologists conduct multi-method assessments, which include gathering information from different informants, in different settings, and using a variety of procedures, such as interviews, observations, behavior rating scales, and standardized and non-standardized tests with excellent psychometric properties. Moreover, school psychologists are conducting assessment for purposes other than eligibility determination, such as screening and progress monitoring. In sum, the current test use and assessment practices of school psychologists exemplifies the evidence-based practice movement, which aims to promote the delivery of services with demonstrated research support to improve the quality of professional services (e.g., diagnosis, evaluation, and intervention) delivered to children and youth, families, and schools (Gross, Farmer, & Ochs, 2018; Kratochwill, 2007).

References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Publishing.
- Balow, I. H., Farr, R. C., & Hogan, T. P. (1992). *Metropolitan achievement test*. 7. San Antonio, TX: Psychological Corporation.
- Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). San Antonio, TX: Pearson.
- Beck, J. S., Beck, A. T., Jolly, J. B., & Steer, R. A. (2005). *Manual for the Beck youth inventories, Second edition for children and adolescents*. San Antonio, TX: Pearson.
- Beery, K. E., Buktenica, N. A., & Beery, N. A. (2010). *The Beery-Buktenica developmental test of visual-motor integration* (6th ed.). New York, NY: Pearson.
- Bender, L. (1946). *Instructions for the use of visual motor gestalt test*. Alexandria, VA: The American Orthopsychiatric Association.
- Benson, N., Kranzler, J. H., & Floyd, R. G. (2016). Examining the integrity of measurement of cognitive abilities in the prediction of achievement: Comparisons and contrasts across variables from higher-order and bifactor models. *Journal of School Psychology, 58*, 1–19. <https://doi.org/10.1080/21683603.2016.116759>.
- Bracken, B. A. (2006). *Bracken basic concept scale-receptive* (3rd ed.). San Antonio, TX: Harcourt Assessment.
- Bracken, B. A., & McCallum, R. S. (2016). *Universal nonverbal intelligence test* (2nd ed.). Austin, TX: PRO-ED.
- Brannigan, G. G., & Decker, S. L. (2003). *Bender visual-motor gestalt test* (2nd ed.). Itasca, IL: Riverside.
- Briesch, A. M., Volpe, R. J., & Floyd, R. G. (2018). *School-based observation: A practical guide to assessing student behavior*. New York, NY: Guilford Press.
- Buck, J. N. (1964). *The house-tree-person*. Beverly Hill, CA: Western Psychological Services.
- Canivez, G. L., & Watkins, M. W. (2016). Review of the Wechsler Intelligence Scale for Children-Fifth Edition: Critique, commentary, and independent analyses. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligence testing with the WISC-V* (pp. 683–702). Hoboken, NJ: Wiley.
- Conners, C. K. (1997). *Conners rating scales*. Revised Toronto, ON: Multi-Health Systems.
- Conners, C. K. (2008). *Conners rating scales* (3rd ed.). Toronto, ON, Canada: Multi-Health Systems.
- Connolly, A. J. (2007). *Keymath-3 diagnostic assessment*. NJ: Upper Saddle Creek.
- Daly, E. J., III, Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26*, 554–574.
- Dean, R. S., & Woodcock, R. W. (2003). *Dean-Woodcock sensory-motor battery*. Itasca, IL: Riverside.
- Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics, 43*, 49–93. <https://doi.org/10.2307/2986113>.
- Drummond, T. (1994). *The student risk screening scale*. Grants Pass, OR: Josephine County Mental Health Program.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd ed.). Bloomington, IN: Pearson.
- Elliot, C. D. (2007). *Differential ability scales* (2nd ed.). San Antonio, TX: Harcourt Assessment.
- Eysenbach, G. (2004). Improving the quality of web surveys: The Checklist for Reporting Results of Internet ESURveys (CHERRIES). *Journal of Medical Internet Research, 29*(6(3)), e34. <https://doi.org/10.2196/jmir.6.6.e34>.

- Fagan, T. K., & Wise, P. S. (2007). *School psychology: Past, present, and future*. Bethesda, MD: National Association of School Psychologists.
- Flanagan, D., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). New York, NY: Wiley.
- Flanagan, D. P., & Harrison, P. (Eds.). (2012). *Contemporary intellectual assessment* (3rd ed.). New York, NY: Guilford Press.
- Floyd, R. G., Reynolds, M. R., Farmer, R. L., & Kranzler, J. H. (2013). Are the general factors from different child and adolescent intelligence tests the same? Results from a five-sample, six-test analysis. *School Psychology Review, 42*, 383–401.
- Floyd, R. G., Shands, E. I., Phillips, J., Autry, B., Mosteller, J., Alfonso, V., ... Irby, S. M. (2015). A systematic review and evaluation of the technical characteristics of adaptive behavior scales. *Journal of Applied School Psychology, 31*, 83–113. <https://doi.org/10.1080/15377903.2014.979.384>.
- Gilliam, J. E. (2013). *Gilliam autism rating scale* (3rd ed.). Austin, TX: PRO-ED.
- Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2015). *The behavior rating inventory of executive function* (2nd ed.). Odessa, FL: Psychological Assessment Resources.
- Goh, D. S., Teslow, C. J., & Fuller, G. B. (1981). The practice of psychological assessment among school psychologists. *Professional Psychology, 12*, 696–706. <https://doi.org/10.1037/0735-7028.12.6.696>.
- Goldstein, S., & Naglieri, J. A. (2009). *Autism spectrum rating scales*. Toronto, ON: Multi-Health Systems.
- Gross, T. J., Farmer, R. L., & Ochs, S. E. (2018). Evidence-based assessment: Best practices, customary practices, and recommendations for field-based assessment. *Contemporary School Psychology*. <https://doi.org/10.1007/s40688-018-0186-x>.
- Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity*. New York, NY: Harcourt, Brace & World.
- Harrison, P. L., & Oakland, T. (2015). *Adaptive behavior assessment system* (3rd ed.). New York, NY: Pearson.
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin card sorting test*. Odessa, FL: Psychological Assessment Resources.
- Hojnoski, R. L., Morrison, R., Brown, M., & Matthews, W. J. (2006). Projective test use among school psychologists: A survey and critique. *Journal of Psychoeducational Assessment, 24*, 145–159. <https://doi.org/10.1177/0734282906287828>.
- Houghton Mifflin Harcourt (2011). *Scholastic reading inventory*. San Antonio, TX: Harcourt Assessment.
- Hutton, J. B., Dubes, R., & Muir, S. (1992). Assessment practices of school psychologists: Ten years later. *School Psychology Review, 21*, 271–284.
- Individuals with Disabilities Education Act (2004). *Pub. L. No. 108–446*.
- Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (Eds.). (2016). *Handbook of response to intervention: The science and practice of multi-tiered systems of support*. New York, NY: Springer.
- Jimerson, S. R., Skokut, M., Cardenas, S., Malone, H., & Stewart, K. (2008). Where in the world is school psychology? Examining evidence of school psychology around the globe. *School Psychology International, 29*, 131–144. <https://doi.org/10.1177/0143034308090056>.
- Johnson, R. L., & Morgan, G. B. (2016). *Survey scales: A guide to development, analysis, and reporting*. New York: Guilford Press.
- Kamphaus, R. W., & Pleiss, K. L. (1991). Draw-a-person techniques: Tests in search of a construct. *Journal of School Psychology, 29*, 395–401. [https://doi.org/10.1016/0022-4405\(91\)90026-n](https://doi.org/10.1016/0022-4405(91)90026-n).
- Kamphaus, R. W., & Reynolds, C. R. (2015). *Behavior Assessment System for Children – Third Edition (BASC-3): System*. Minneapolis, MN: Pearson Assessment.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children* (2nd ed.). Circle Pines, MN: American Guidance Service.
- Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman test of educational achievement* (3rd ed.). New York, NY: Pearson.
- Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., ... Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*, 980–988. <https://doi.org/10.1097/00004583-199707000-00021>.
- Kilgus, S. P., Eklund, K., von der Embse, N. P., Taylor, C. N., & Sims, W. A. (2016). Psychometric defensibility of the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) Teacher Rating Scale and multiple gating procedure within elementary and middle school samples. *Journal of School Psychology, 58*, 21–39. <https://doi.org/10.06/j/jsp.2016.07.001>.
- Kirk, S. A., McCarthy, J. J., & Kirk, W. D. (1968). *Illinois test of psycholinguistic abilities: Revised edition*. Urbana, IL: University of Illinois Press.
- Kolowich, L. (2018, April 4). The best time to post on Instagram, Facebook, Twitter, LinkedIn, Pinterest, and Google+ [Blog post]. Retrieved from <https://blog.hubspot.com/marketing/best-times-post-pin-tweet-social-media-infographic#WZ8kDBPK8YQ.twitter>.
- Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: A developmental neuropsychological assessment* (2nd ed.). San Antonio, TX: Psychological Corporation.
- Kovacs, M. (2010). *The children's depression inventory* (2nd ed.). Toronto, ON: Multi-Health Systems.
- Kranzler, J. H., Benson, N., & Floyd, R. G. (2016). Intellectual assessment of children and youth in the United States of America: Past, present, and future. *International Journal of School and Educational Psychology, 4*, 276–282. <https://doi.org/10.1080/21683603.2016.1166759>.
- Kranzler, J. H., & Floyd, R. G. (2013). *Assessing intelligence in children and adolescents: A practical guide*. New York, NY: Guilford Press.
- Kratochwill, T. (2007). Preparing psychologists for evidence-based school practice: Lessons learned and challenges ahead. *American Psychologist, 62*, 829–843. <https://doi.org/10.1037/0003-066x.62.8.829>.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1*, 27–66. <https://doi.org/10.1111/1529-1006.002>.
- Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90*, 1112–1121.
- Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism diagnostic observation schedule* (2nd ed.). Los Angeles, CA: Western Psychological Services.
- Mardell, C., & Goldenberg, D. S. (2011). *Developmental indicators for the assessment of learning* (4th ed.). New York, NY: Pearson.
- Markwardt, F. C. (1989). *Peabody individual achievement test* (Revised) Circle Pines, MN: American Guidance Service.
- Morgan, C. D., & Murray, H. A. (1935). A method for investigating phantasies. The Thematic Apperception Test. *Archives of Neurology and Psychiatry, 34*, 289–306.
- National Association of School Psychologists (2010). *Model for comprehensive and integrated school psychological services*. Bethesda, MD: National Association of School Psychologists.
- National Center for Education Statistics (2018, May 29). Children and Youth With Disabilities. Retrieved from https://nces.ed.gov/programs/coe/indicator_cgg.asp.
- National Center on Intensive Intervention (2018). *Academic screening tools chart*. Department of Education, Office of Special Education Programs, National Center on Intensive Intervention: Washington, DC: U.S Retrieved from <http://www.intensiveintervention.org/chart/academic-screening>.
- Newborg, J. (2005). *Battelle developmental inventory* (2nd ed.). Itasca, IL: Riverside.
- Reitan, R. M. (1979). *Manual for administration of neuropsychological test batteries for adults and children*. Tucson, AZ: Author.
- Reschly, D. J., Genshaft, L., & Binder, M. S. (1987). *The 1986 NASP survey: Comparison of practitioners, NASP leadership, and university faculty on key issues*. Washington, DC: National Association of School Psychologists.
- Reynolds, C. R., & Kamphaus, R. W. (2015a). *Behavioral assessment system for children* (3rd ed.). New York, NY: Pearson.
- Reynolds, C. R., & Kamphaus, R. W. (2015b). *Reynolds intellectual assessment scales* (2nd ed.). Lutz, FL: Psychological Assessment Resources.
- Reynolds, C. R., & Richmond, B. O. (2008). *Revised children's manifest anxiety scale* (2nd ed.). Los Angeles, CA: Western Psychological Services.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581–592. <https://doi.org/10.2307/2335739>.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177. <https://doi.org/10.1037/1082-989x.7.2.147>.
- Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan, & E. M. McDonough (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 73–162). (4th ed.). New York, NY: Guilford.
- Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *The childhood autism rating scale* (2nd ed.). Toronto, ON: Multi-Health Systems.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV tests of cognitive abilities*. Rolling Meadows, IL: Riverside.
- Shapiro, E. S. (2004). *Academic skills problems workbook* (rev. ed.). New York, NY: Guilford Press.
- Shapiro, E. S., Angello, L. M., & Eckert, T. L. (2004). Has Curriculum-Based Assessment become a staple of school psychology practice? An update and extension of knowledge, use and attitudes from 1990 to 2000. *School Psychology Review, 33*, 249–257.

- Sheslow, D., & Adams, W. (2003). *Wide range assessment of memory and learning* (2nd ed.). New York, NY: Pearson.
- Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools, 51*, 1031–1045. <https://doi.org/10.1002/pits.21802>.
- Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidance Service.
- Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). *Vineland adaptive behavior scales* (3rd ed.). New York, NY: Pearson.
- Splett, J. W., Fowler, J., Weist, M. D., McDaniel, H., & Dvorsky, M. (2013). The critical role of school psychology in the school mental health movement. *Psychology in the Schools, 50*, 245–258. <https://doi.org/10.1002/pits.21677>.
- Stinnett, T. A., Harvey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment, 12*, 331–350.
- The Psychological Corporation (1988). *Stanford achievement test*. San Antonio, TX: Author.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford–Binet intelligence scale* (4th ed.). Chicago, IL: Riverside.
- VandenBos, G. R. (2006). *APA dictionary of psychology*. Washington, DC: American Psychological Association <https://doi.org/10.1037/14646-000>.
- Visser, L., Ruiter, S. A. J., van der Meulen, B. F., Ruijsenaars, W. A. J. J. M., & Timmerman, M. E. (2012). A review of standardized developmental assessment instruments for young children and their applicability for children with special needs. *Journal of Cognitive Education and Psychology, 11*, 102–127. <https://doi.org/10.1891/1945-8959.11.2.102>.
- Volpe, R. J., & Fabiano, G. A. (2013). *Daily behavior report cards: An evidence-based system of assessment and intervention*. New York, NY: The Guilford Press.
- Voress, J. K., & Maddox, T. (2012). *Developmental assessment of young children* (2nd ed.). Austin, TX: PRO-ED.
- Wagner, R., Torgesen, J., Rashotte, C., & Pearson, N. A. (2013). *Comprehensive test of phonological processing* (2nd ed.). New York, NY: Pearson.
- Walcott, C. M., & Hyson, D. (2018). *Results from the NASP 2015 membership survey, part one: Demographics and employment conditions [Research report]*. Bethesda, MD: National Association of School Psychologists.
- Wechsler, D. (1974). *Wechsler intelligence scale for children—Revised*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1991). *Wechsler intelligence scale for children* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2009). *Wechsler individual achievement test* (3rd ed.). New York, NY: Pearson.
- Wechsler, D. (2014). *Wechsler intelligence scale for children* (5th ed.). New York, NY: Pearson.
- Wechsler, D. (2015). *Wechsler intelligence scale for children, fifth edition integrated*. New York, NY: Pearson.
- Whitcomb, S. A. (2017). *Behavioral, social, and emotional assessment of children and adolescents* (5th ed.). New York, NY: Routledge.
- Wiederholt, J. L., & Bryant, B. (2012). *Gray oral reading test* (5th ed.). Austin, TX: PRO-ED.
- Wilkinson, G. S. (1993). *Wide range achievement test* (3rd ed.). Wilmington, DE: Wide Range.
- Wilkinson, L. A. (2010). *A best practice guide to assessment and intervention for autism and Asperger's syndrome in schools*. London, UK: Jessica Kingsley.
- Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review, 25*, 9–23.
- Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson psycho-educational battery, Revised*. Itasca, IL: Riverside.