Morash, V. & McKerracher, A. (2016). Low reliability of sighted-normed verbal assessment scores when administered to children with visual impairments. *Psychological Assessment*, (accepted).

Abstract

The most common and advocated assessment approach when a child cannot access visual materials is to use the verbal subscales of a test the psychologist already has and is familiar with. However, previous research indicates that children with visual impairments experience atypical verbal development. This raises the question of whether verbal subscale scores retain their reliability and interpretation validity when given to children with visual impairments. To answer this question, we administered a vocabulary subscale from a common intelligence test along with several non-verbal subscales to 15 early-blind adolescents (onset $\leq$ 2 years). Reliability of only the vocabulary test scores was insufficient for high-stakes testing. This finding points to the broader issue of difficulties in assessing populations of exceptional children who experience atypical development trajectories, possibly making their assessment with common tests inappropriate.

Keywords: Assessment, Intelligence, Verbal, Visual Impairment, Internal Consistency, Reliability

Low Reliability of Sighted-Normed Verbal Assessment Scores when Administered to Children with Visual Impairments

When a child has a visual impairment, the child is at an increased risk for learning disabilities, developmental disorders, autism spectrum disorders (ASD), and mild traumatic brain injuries, making psychological assessment particularly important (Dial & Dial, 2010). Due to the movement of educating children with special needs in mainstream schools, and no longer concentrating children with visual impairments at specialist schools for the blind, the assessment of children with visual impairments is often conducted by psychologists who serve the general public (Bauman & Kropf, 1979). When faced with assessing a child who cannot access visual materials, these psychologists are likely to turn to verbal scales they have available and are known to produce reliable scores when assessing sighted children (Miller & Skillman, 2003).

Psychological tests are used to estimate a child's overall and specific intellectual skills relative to his/her peers. This information contributes to decisions about whether the child has a learning disability (Decker, Hale, Flanigan, 2013), an intellectual disability (Bergeron, Floyd, & Shands, 2008), or should be referred for gifted services (Gross, 2004; Pfeiffer, 2012). Psychological tests that are used in these contexts are considered "high stakes" because each is one of only a handful that is used to make important decisions about the child's educational and clinical services. In contrast, low stakes assessments, such as classroom exams, are just one of many sources, e.g., other exams and homework, used to determine the child's grade or placement.

A common approach to the psychological assessment of children with visual impairments is to administer only the verbal subtests from a more comprehensive intelligence test (Crepeau-Hobson & Vujeva, 2012; Miller & Skillman, 2003; Nelson, Dial, & Joyce, 2002). The most popular psychological tests are the Wechsler intelligence scales, e.g., the Wechsler Intelligence Scale for Children (WISC) and the Wechsler Adult Intelligence Scale (WAIS; Kamphaus, Petoskey, & Rowe, 2000; Kaplan & Saccuzzo, 2008). The verbal subtests from Wechsler intelligence scales are the most common (Miller & Skillman, 2003; Nelson et al., 2002) and recommended tests for the assessment of children with visual impairments, although there are some concerns that use of these subtests, alone, neglects nonverbal abilities (Bauman & Kropf, 1979; Dekker, 1993; Gutterman, Ward, & Genshaft, 1985; Lund, Miller, & Ganz, 2014; Miller & Skillman, 2003). Verbal subtests are used for intellectual assessment, along with observational methods, neuropsychological assessment, visual assessment, and assessment of skills, by psychologists tasked with the comprehensive evaluation of children with visual impairments (Bradley-Johnson & Ekstrom, 1999).

Many researchers have investigated whether scores obtained using sighted-normed verbal subtests produce inflated or deflated results when applied to children with visual impairments. Children who are visually impaired tend to score below their sighted peers on comprehension, similarity, and vocabulary subtests (Dekker, 1993; Tillman & Bashaw, 1968; Vander Kolk, 1977), which is attributed to fewer opportunities to benefit from traditional classroom education, lack of visual daily experiences, and the inability to use visual cues in the assessment environment, e.g., when asked to generate a list of objects (Dial & Dial, 2010; Nelson et al., 2002; Vander Kolk, 1977; Wyver et al., 1999). In contrast, children with visual impairments score consistently higher than their sighted peers on tests of verbal working memory, such as the digit span subtest (Dekker, 1993; Hull & Mason, 1995; Tillman & Bashaw, 1968), presumably due to daily practice using verbal working memory strategies, such as counting steps, to complete tasks without vision (Hull & Mason, 1995). Despite deviations in specific subtests, the use of sighted-normed verbal subscales is generally considered valid and valuable, although it is recommended that results should be interpreted cautiously due to these norming issues (Bradley-Johnson & Ekstrom, 1999; Miller & Skillman, 2003). Comparison to sighted norms, even if not perfect, is believed to provide valuable insights into a child's functioning relative to sighted standards (Dial & Dial, 2010).

So far, concerns about verbal subtests administered to children with visual impairments have focused on the use of sighted norms resulting in inflated or deflated scores. However, it is also reasonable to question whether verbal tests developed for sighted children produce scores that are reliable and interpretations valid (beyond norm discrepancies) when applied to children with visual impairments. Certainly, modification of assessments, such as using tactile graphics in place of visual graphics on achievement tests, is known to alter item functioning for children who are blind (Bennett, Rock, & Novatkoski, 1989; Morash & McKerracher, 2014). Although it may seem unlikely that verbal items would function differently for children with visual impairments, because the assessment can be administered without modification, children with visual impairments exhibit atypical conceptual and semantic development (Pring, 1988; Pring, Freestone, & Katan, 1990), which may affect their performance on verbal tests. Children with visual impairments also demonstrate specific deficits in understanding visual concepts, such as color, which they have learned through language and not direct experience (Wyver et al., 1999); and paradoxically receive less verbal information from their caregivers describing objects, people, and events than sighted children (Andersen, Dunlea, & Kekelis, 1993).

Due to altered language functioning, verbal subscales, especially those that tap semantic knowledge such as vocabulary tests, may produce scores with low reliability when given to children with visual impairments. Reliability limits validity, the extent to which test scores reflect what the test purports to measure (Brown, 1996), as reliability constrains how strongly the test scores can correlate with another, e.g., external, measure (John & Soto, 2007). Without sufficient reliability, scores are highly susceptible to random error, causing deviations from true scores that are too low or high, and reducing their usefulness for analysis, interpretation, and clinical purposes (Henson, 2001).

For low stakes assessments and research purposes, reliability between 0.70 and 0.90 is considered suitable, while for high-stakes assessments used for important clinical and/or education decisions, reliability should be 0.90 or higher (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999; Hopkins, 1998; Kaplan & Saccuzzo, 2008, p. 125; Nunnally, 1967, p. 226; Nunnally & Bernstein, 1994, p. 265; Oosterhof, 201l). Given that practitioners are likely to use verbal subtests in high-stakes assessments of children with visual impairments, the higher levels of reliability should be met. Here, we test whether the reliability of sighted-normed vocabulary subtest scores is sufficient for high-stakes assessment of children with visual impairments.

Methods

Participants

Participants were chosen to produce a homogenous sample in regards to age, level of visual impairment, age of onset, and additional disabilities. Although this is not a representative sample of the heterogeneous population of children with visual impairments, these variables can affect test functioning (Dial & Dial, 2010). Therefore, we wanted to control these variables in the current study to produce clear findings, while future studies may examine other subpopulations of children with visual impairments or the heterogeneous population at large. Enrolling participants meeting the inclusion criteria required a substantial effort due to the low incidence of early blindness without multiple disabilities.

Fifteen adolescents, mean age 15.71 years (SD=2.08), were recruited who were legally blind (visual acuity worse than 20/200 in both eyes) with early onset (n=13 congenital, n=1 onset at 0.75 years, n=1 onset at 2 years), had no form perception, used braille (not large print or magnification) to access printed materials, and had no additional disabilities. All participants were native English speakers and planned to receive their high-school diplomas by age 21 (the typical age cutoff for attending American high schools). Causes of visual impairment included Aniridia and Glaucoma (n=1), Bilateral Anophthalmia (n=2), Bilateral Microphthalmia (n=1), Familial Exudative Vitreoretinopathy (FEVR; n=1), Leber Congenital Amaurosis (LCA, n=2), Norrie Disease (n=1), Optic Nerve Hypoplasia (OH, n=1), Retinal Detachment (n=1), Retinopathy of Prematurity (RoP, n=3), and unknown (n=2). The University of California, Berkeley's Institutional Review Board approved the study's procedures, and participants and their parents (when participants were under 18) provided informed consent to participate in this research.

Materials & Procedures

Participants completed a vocabulary and other tests in random order. Non-vocabulary tests were included to provide information on how score reliability varied in our sample.

Vocabulary. The Vocabulary subtest was selected from the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), which is regarded as an exceptionally well standardized and widely used short form assessment (e.g., Minshew, Turner, Goldstein, 2005). The WASI is designated for use with individuals aged 6 to 89 years old, and its subtests are similar to their WISC-III and WAIS-III counterparts (Wechsler, 1999). Average internal consistency reliability for the WASI Vocabulary subtest scores is reported to be 0.90 for the ages of our participants (Wechsler, 1999).

Rapid Symbolic Naming. Rapid symbolic naming was assessed using two subtests, Rapid Digit Naming and Rapid Letter Naming, from the Comprehensive Test of Phonological Processing (CTOPP), which is designated for use with individuals aged 7 to 24 years old (Wagner, Torgesen, Raschotte, & Pro-Ed, 1999). These subtests measure the participant's ability to efficiently retrieve phonological information form long-term memory and execute a sequence of operations. Participants are scored based on how quickly they can read aloud a list of printed digits or letters. In the current study, letter and digit naming tests were administered as alternate forms. The average reliability of the Rapid Letter Naming subtest scores for sighted individuals is reported to be .84, and for the Rapid Digit Naming subtest scores .88, for the ages of our participants (Wagner et al., 1999). To make these tests accessible to our participants, we replaced the original printed materials with braille equivalents.

Oral Reading Fluency. Reading fluency was measured with the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Next, which is designated for grades K-6 (Good, Kaminski, Sopris West, 2003). Participants are scored on the number of words correctly read aloud from a connected-text passage in 60 seconds. This assessment is available in both uncontracted and contracted braille, although norms are only available for the print version. We administered two measures from the grade 6 (the oldest grade available) DIBELS Oral Reading Fluency (DORF) assessment in contracted braille, benchmarks 1.1 and 1.2. The single-passage parallel-forms reliability associated with the print DORF scores for grades 1-6 is 0.91 (Good et al., 2003). Although the participants in the current study were all above grade 6, their average performance was still below the sighted grade-6 benchmark (120 words per minute). In fact, two participants were below the benchmark and ten more were well below the benchmark (95 words per minute) according to sighted norms (Good et al., 2003). This was likely due to braille reading fluency being significantly lower than visual print reading fluency (Wetzel & Knowlton, 2000).

Blind Learning Aptitude Test. Developed specifically to test the nonverbal reasoning of children with visual impairments, the Blind Learning Aptitude Test (BLAT) consists of 49 tactile-graphics items, which includes pattern completion, figure completion, and matrix reasoning tasks. This assessment has not been updated since initial development and norming in 1969, when internal consistency was reported to be 0.93 for 961 blind children aged 6-20 years old (Newland, 1969).

Auditory Simple Reaction Time. Measures of participants' simple reaction times were collected using procedures similar to other research (e.g., Vernon, 1983) implemented in Matlab using the Psychophysics Toolbox extensions (Brainard, 1997). Participants were asked to press a keyboard button with the preferred finger of their preferred hand as soon as possible after hearing a 200 ms tone. Timing between the tones varied randomly according to a uniform distribution between two and five seconds. Participants executed 5 practice trails and 30 test trials.

Analyses

Score reliability was estimated using internal consistency: Cronbach's alpha (α), split-half reliability (even and odd items), and KR-20; and parallel-forms, depending on the nature of the test (Henson, 2001). Cronbach's α and split-half reliability were applied to all non-timed tests (Vocabulary, Nonverbal Reasoning, and Reaction Time). Split-half reliability was corrected for whole (equal length) reliability using the Spearman-Brown correction, and may be considered the most appropriate reliability measure for Vocabulary and Nonverbal Reasoning, which had items ordered by difficulty. KR-20 is a more specific form of Cronbach's α, applied only when items receive binary (right/wrong) scores (Nonverbal Reasoning). Tests administered as alternate forms (Rapid Automatic Naming and Oral Reading Fluency) had scores assessed using parallel-forms reliability.

We computed 95% confidence intervals for reliability estimates to provide insight into the larger population from which our sample was drawn. Confidence intervals were estimated by nonparametric bootstrap (Monte Carlo case resampling of participants), with 100,000 iterations and a sample size equal to the number of participants. Each confidence interval reflected the uncertainty associated with our sampling method, which selected participants from a population of children meeting our inclusion criteria (see Participants). In particular, if we were able to repeat our study many times, in each instance selecting a new participant sample, the 95% confidence interval would contain the population's score reliability 95% of the time. Equivalently, across all studies (not just replications of the current one), a 95% confidence interval contains the true parameter of interest (in this report score reliability) 95% of the time (Wasserman, 2004).

There is a correspondence between confidence intervals and hypothesis testing, which allowed us to compare the current estimated reliabilities to the recommended 0.90 value (taken as the null hypothesis). All reliability values contained in a 95% confidence interval are considered plausible and cannot be rejected. Therefore, if the null-hypothesis value fell within the 95% confidence interval, then the null hypothesis was not rejected (at $\alpha = 0.05$). If the null-hypothesis value fell outside the 95% confidence interval, then the null hypothesis was rejected (at α = 0.05). Thereby, we compared reliability findings to the recommended 0.90 value using confidence intervals.

Results & Discussion

Table 1. *Test Scores and Internal Consistency Results*

	Scores Mean (SD)	Internal Consistency Estimate (95% CI)
Vocabulary (WASI)	*Overall Raw Score:* 56.47 (8.18) pts *Overall T-Score:* [a] 58.47 (10.65) pts	*Cronbach's α*: 0.85 (0.74, 0.89) *Split-Half Reliability*: 0.72 (0.41, 0.89)
Rapid Automatic Naming (CTOPP)	*Overall Speed:* 18.00 (3.14) s *Letter:* [b] 17.80 (3.14) s *Digit:* [b] 18.00 (1.69) s	*Parallel Forms:* 0.90 (0.76, 0.96)
Oral Reading Fluency (DIBELS)	*Overall:* 98 (35) words *Benchmark 1.1:* [c] 102 (32) words *Benchmark 1.2:* [c] 95 (38) words	*Parallel Forms*: 0.98 (0.95, 0.99)
Nonverbal Reasoning (BLAT)	*Overall Score:* 34.73 (9.39) pts *Item Accuracy:* 70.88 (45.46) %	*Cronbach's α*: 0.92 (0.81, 0.95) *KR-20*: 0.92 (0.83, 0.95) *Split-Half Reliability*: 0.81 (0.57, 0.94)
Reaction Time	374 (237) ms	*Cronbach's α*: 0.98 (0.93, 0.98) *Split-Half Reliability*: 1.00 (0.90, 1.00)

[a] Caution advised, T-score calculation relies on sighted norms.
[b] paired $t(14) = 0.76$, $p = 0.458$ n.s.
[c] paired $t(14) = 2.39$, $p = 0.031$ *

Test performance and score reliability results are shown in Table 1. No participant performed at floor or ceiling (receiving close to the lowest or highest scores) on any of the tests.

The reliabilities of vocabulary test scores were significantly lower (based on 95% confidence intervals) than the cutoff for high-stakes assessment (0.90). This is particularly problematic, because a vocabulary subtest is likely to be one of a limited number of tests used during high-stakes assessments of children with visual impairments, because vocabulary subtests are found in common intelligence tests and can be used without adaptation. In practice, practitioners are likely to lean on standardized assessments because they are unfamiliar with visual impairment, which is both relatively low-incidence (occurring in 0.3 per 1000 births; Steinkuller et al., 1999) and extremely heterogeneous (Scholl, 1987). While it is recommended that assessors rely on additional sources of information when testing children with visual impairments, such as behavioral observation to assess a student's day-to-day performance and a functional behavioral assessment (Lund et al., 2014), mainstream teachers and psychologists who do not specialize in visual impairment are unlikely to recognize whether behaviors of children with visual impairments are typical/atypical or functional/maladaptive. While the other test scores we examined had adequate reliability, it would be unusual for general practitioners to have access to these materials, especially the braille and tactile-graphics versions of the CTOPP, DIBELS, and BLAT.

We do not believe that the low reliability of vocabulary test scores was due to unusual characteristics of our participants, because the other tests we investigated had high ($\geq$0.90) score reliabilities (none were significantly below 0.90). Furthermore, while lower score variance can result in a lower coefficient alpha (Rodriguez & Maeda, 2006), the T-score variance of blind participants (*SD* = 10.65, Table 1) was similar to that reported for similarly aged sighted participants (*SD* = 9.95, Wechsler, 1999). Vocabulary may be particularly susceptible to the atypical language development of children with visual impairments (Pring, 1988; Pring et al., 1990). In particular, the low internal consistency reliability found for vocabulary test scores indicates that the test did not adequately and thoroughly capture the intended single latent trait in children with visual impairments (John & Sato, 2007). This may be because there are some words on the vocabulary test that rely on visual experience, and these words function differently for children who do and do not have visual impairments. Alternatively, many non-visual experiences are altered by visual impairment. For example, children with visual impairments may associate the word "lunch," which appeared on the vocabulary subtest we administered, with social anxiety and isolation more so than children who do not have disabilities.

One limitation of the current research is its relatively small sample size. However, we were able to partially address this limitation by calculating confidence intervals, which provide information about the uncertainty associated with our sampling method. The small sample size was due to difficulties finding children who were early blind and had no other disabilities. Children with visual impairments are no longer concentrated at schools for the blind, making them difficult to locate. Furthermore, most children with visual impairments have additional disabilities (Bradley-Johnson & Ekstrom, 1999; Dial & Dial, 2010), which we excluded to obtain a homogenous sample. While these issues complicate the development and evaluation of assessments for children with visual impairments, they also motivate such efforts. The need for quality assessments for children with visual impairments is increased by the fact that children with visual impairments are at an increased risk for additional disabilities and disorders that may affect their learning (Bradley-Johnson & Ekstrom, 1999; Dial & Dial, 2010), and assessment is likely to be carried out by psychologists serving the general public who are not experts in visual impairment. Use of appropriate tests that reflect a child's cognitive capacity and not his/her sensory impairment is required by ethical guidelines (American Psychological Association

[APA], 1992; National Association of School Psychologists [NASP], 2010) and by law (Individuals with Disabilities Education Act [IDEA], 1990, p. 42496).

Unfortunately, psychologists who are mandated by these ethical and legal guidelines to use appropriate assessments for children with visual impairments are in a difficult situation, due to a lack of appropriate scales. While using verbal subtests is the status quo (Crepeau-Hobson & Vujeva, 2012; Miller & Skillman, 2003; Nelson, Dial, & Joyce, 2002), this report exposes problems with this approach. It has long been known that verbal subtests produce inflated or deflated scores when administered to children with visual impairments (Dekker, 1993; Tillman & Bashaw, 1968; Vander Kolk, 1977). Furthermore, the current report indicates that these scales may produce unreliable scores for children with visual impairments. If test scores have low reliability, psychologist cannot trust that results reflect meaningful variation instead of random error (John & Sato, 2007; Henson, 2001), and should not use these tests scores (especially) in high-stakes assessments. It is obvious that there is a serious dearth of psychological assessments that are accessible and developed specifically for children with visual impairments. Not only do such assessments need to have scores with adequate reliability, but their score interpretations need to have established validity among the population of children with visual impairments (which was not addressed in the current report). A test that produces reliable scores and valid inferences with sighted children is not guaranteed to do so with any other population, as these are a function of sample, instrument, administration setting, and purpose (Dawis, 1987; Loevinger, 1957).

In conclusion, our findings are consistent with past research that found that semantic and conceptual knowledge develop differently in individuals with visual impairments than in typically sighted individuals (Pring, 1988; Pring et al., 1990). As we predicted based on these results, we found the scores of verbal subtests from available psychological assessments may not perform adequately with children who have visual impairments. Future work is needed not only to develop assessments for children with visual impairments, but to make these assessments widely available. Our results also beg the question of whether there are other tests performing inadequately with special populations. Practitioners must be extremely cautious or disinclined to use assessments developed for and evaluated with typically-abled children when conducting high-stakes psychological assessments of children from special populations that are likely to experience atypical developmental trajectories.

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47, 1597-1611. http://dx.doi.org/10.1037//0003-066x.47.12.1597

Andersen, E. S., Dunlea, A., & Kekelis, L. S. (1993). The impact of input: Language acquisition in the visually impaired. *First Language*, 13, 23-49. http://dx.doi.org/10.1177/014272379301303703

Bauman, M. K., & Kropf, C. A. (1979). Psychological tests used with blind and visually handicapped persons. *School Psychology Digest*, 8, 257-269.

Bennett, R. E., Rock, D. A., Novatkoski, I. (1989). Differential item functioning on the SAT-M braille edition. *Journal of Educational Measurement*, 26, 67-79. http://dx.doi.org/10.1111/j.1745-3984.1989.tb00319.x

Bradley-Johnson, S., & Ekstrom, R. (1999). Visual impairment. In J. Danoval, C. Frisby, K. Geisinger, J. Scheuneman, & J. Grenier (Eds.), Test interpretation and diversity: Achieving equity in assessment (pp. 271–296). Washington, DC: American Psychological Association.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision,* 10, 433-436. http://dx.doi.org/10.1163/156856897x00357

Bergeron, R., Floyd, R. G., & Shands, E. I. (2008). States' eligibility guidelines for mental retardation: An update and consideration of part scores and unreliability of IQs. *Education and Training in Developmental Disabilities*, 43, 123-131.

Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.

Crepeau-Hobson, F. & Vujeva, H. (2012). Assessment of cognitive ability of students with severe and low-incidence disabilities – Part 1. *Communique*, 41, 12-13.

Dawis, R. V. (1987). Scale construction. *Journal of Counseling Psychology*, 34, 481-489. http://dx.doi.org/10.1037/0022-0167.34.4.481

Decker, S. L., Hale, J. B., & Flanagan, D. P. (2013). Professional practice issues in the assessment of cognitive functioning for educational applications. *Psychology in the Schools*, 50, 300-313. http://dx.doi.org/10.1002/pits.21675

Dekker, R. (1993). Visually impaired children and haptic intelligence test scores: Intelligence Test for Visually Impaired Children (ITVIC). *Developmental Medicine & Child Neurology*, 35, 478-489. http://dx.doi.org/10.1111/j.1469-8749.1993.tb11678.x

Dial, J. G., & Dial, C. L. (2010). Assessing and intervening with visually impaired children and adolescents. In D.C. Miller (ed.) *Best Practices in School Neuropsychology: Guidelines for Effective Practice, Assessment, and Evidence-Based Intervention* (pp. 465-479). Hoboken, NK: John Wiley & Sons, Inc.

Good, R. H., Kaminski, R. A., & Sopris West Inc. (2003). *Dynamic indicators of basic early literacy skills*. Longmont, CO: Sopris West Educational Services.

Gross, M. U. (2003). *Exceptionally gifted children* (2nd ed.). New York: RoutledgeFalmer.

Gutterman, J. E., Ward., M., Genshaft, J. (1985). Correlations of Scores of Low Vision Children on the Perkins-Binet Tests of Intelligence for the Blind, the WISC-R and the WRAT. *Journal of Visual Impairment & Blindness*, 79, 55-58.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, 34, 177-189.

Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed.). Boston: Allyn & Bacon.

Hull, T., & Mason, H. (1995). Performance of blind children on digit-span tests. *Journal of Visual Impairment & Blindness*, 89, 166-169.

Individuals with Disabilities Education Act. (1990). Public Law 101-476, §1401.

John, O. P., & Soto, C. J. (2007). The importance of being valid. *Handbook of Research Methods in Personality Psychology,* 461-494.

Kamphaus, R. W., Petoskey, M. D., & Rowe, E. W. (2000). Current trends in psychological testing of children. *Professional Psychology: Research and Practice*, 31, 155-164. http://dx.doi.org/10.1037//0735-7028.31.2.155

Kaplan, R., & Saccuzzo, D. (2008). *Psychological testing: Principles, applications, and issues* (7th ed.). Belmont, CA: Wadsworth, Cengage Learning.

Lund, E. M., Miller, K. B., & Ganz, J. B. (2014). Access to assessment? Legal and practical issues regarding psychoeducational assessment in children with sensory disabilities. *Journal of Disability Policy Studies*, 25, 135-145. http://dx.doi.org/10.1177/1044207313478661

Loevinger, J. (1957). Objective tests as instruments of psychological theory: Monograph supplement 9. *Psychological Reports*, 3, 635-694. http://dx.doi.org/10.2466/pr0.1957.3.3.635

Miller, J. C., & Skillman, G. D. (2003). Assessors' satisfaction with measures of cognitive ability applied to persons with visual impairments. *Journal of Visual Impairment & Blindness*, 97, 769-774.

Minshew, N. J., Turner, C. A., & Goldstein, G. (2005). The application of short forms of the Wechsler intelligence scales in adults and children with high functioning autism. *Journal of Autism and Developmental Disorders*, 35, 45–52. http://dx.doi.org/10.1007/s10803-004-1030-x

Morash, V. S. & McKerracher, A. (2014). The relationship between tactile graphics and mathematics for students with visual impairments. *Terra Haptica*, 4, 13-22.

National Association of School Psychologists (2010). *Principles for professional ethics*. Bethesda, MD: Author.

Nelson, P. A., Dial, J. G., & Joyce, A. (2002). Validation of the cognitive test for the blind as an assessment of intellectual functioning. *Rehabilitation Psychology*, 47, 184–193. http://dx.doi.org/10.1037/0090-5550.47.2.184

Newland, T.E. (1969). The Blind Learning Aptitude Test. Report, USOE Grant No. OEG-3-6-061928. Urbana, IL: University of Illinois.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

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

Oosterhof, A. (2001). *Classroom applications of educational measurement* (3rd ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

Pfeiffer, S. I. (2011). Current perspectives on the identification and assessment of gifted students. *Journal of Psychoeducational Assessm*ent, 30, 3-9. http://dx.doi.org/10.1177/0734282911428192

Pring, L. (1988). The "reverse-generation" effect: A comparison of memory performance between blind and sighted children. *British Journal of Psychology*, 79, 387-400. http://dx.doi.org/10.1111/j.2044-8295.1988.tb02297.x

Pring, L. Freistone, S. E., & Katan, S. A. (1990). Recalling pictures and words: Reversing the generation effect. *Current Psychological Research & Reviews*, 9, 35-45. http://dx.doi.org/10.1007/bf02686766

Rodriguez, M.C. & Maeda, Y. (2006). Meta-analysis of coefficient alpha. Psychological Methods, 11, 306-322. http://dx.doi.org/10.1037/1082-989X.11.3.306

Scholl, G. T. (1987). Appropriate education for visually handicapped students. *Teaching Exceptional Children,* 19, 32-36.

Steinkuller, P. G., Du, L., Gilbert, C., Foster, A., Collins, M. L., & Coats, D. K. (1999). Childhood blindness. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 3, 26-32.

Tillman, M. H., & Bashaw, W. L. (1968). Multivariate analysis of the WISC scales for blind and sighted children. *Psychological Reports*, 23, 523-526. http://dx.doi.org/10.2466/pr0.1968.23.2.523

Vander Kolk, C. J. (1977). Intelligence testing for visually impaired persons. *Journal of Visual Impairment & Blindness*, 71, 158-163.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence*, 7, 53-70. http://dx.doi.org/10.1016/0160-2896(83)90006-5

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pro-Ed (Firm). (1999). *CTOPP: Comprehensive test of phonological processing*. Austin, Tex: Pro-Ed.

Wasserman, L. (2004). *All of statistics: A concise course in statistical inference.* New York: Springer.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence WASI: Manual*. San Antonio: Pearson/PsychCorpl.

Wetzel, R., & Knowlton, M. (2000). A comparison of print and braille reading rates on three reading tasks. *Journal of Visual Impairment & Blindness*, 94, 146-154.

Wyver, S. R., Markham, R., & Hlavacek, S. (1999). Visual items in tests of intelligence for children. *Journal of Visual Impairment & Blindness*, 93, 663-665.