

Exploring Assessment Processes in Specialized Schools for Students Who Are Visually Impaired

Cheryl Kamei Hannan

Abstract: In this qualitative study, various professionals in specialized schools for students who are visually impaired provided information on assessment tools; how information was used to plan Individualized Education Programs; and their opinions on the reliability, validity, and usefulness of various measurements. The implications of the findings for policy guidelines and high-stakes decisions are explored.

The author thanks Jane Erin and Alberto Arenas of the University of Arizona and Sharon Sacks of California School for the Blind for their help in preparing this manuscript.

Although educators contend that assessment is the backbone of educational programming and a necessary tool in planning individualized instruction, assessment is useful only if it yields meaningful measurements (Genshaft, Dare, & O'Malley, 1980). According to the policy of the No Child Left Behind (NCLB) Act (2001), schools are expected to demonstrate adequate yearly progress, and all children are being challenged to maintain performance at their grade levels. Standardized tests are being used to measure individual performance and adequate yearly progress. Because accountability is measured by these test scores, educators must consider how appropriate they are for students who are visually impaired (that is, those who are blind or have low vision) (Bolt & Thurlow, 2004; Ekstrom, 1998; Reid 1998). Similarly, if educators are required to provide evidence of students' annual progress on Individualized Education Programs

(IEPs), then the validity of all assessment tools and accommodations that are used to measure progress must be measured.

The purpose of this qualitative study was to examine the academic assessment process that is used in specialized schools for students who are visually impaired. I particularly investigated what tools were being used and how the data shaped the IEPs. During the interviews, professionals gave their opinions about the reliability, validity, and usefulness of various standardized and nonstandardized tests. Such information will be beneficial to educators and administrators who need to choose assessment tools for students who are visually impaired.

Relevant literature

BACKGROUND

Despite the mandates of the Individuals with Disabilities Education Improvement Act (IDEIA) of 1997, which requires special education students to be included in state testing, students with visual impairments have historically been exempt from these tests (Ekstrom, 1998). In addition, state tests are not always available or accessible in alternative media, such as braille or large print (Bolt & Thurlow, 2004), thereby excluding many students who need alternate media from test-taking procedures. The practice of excluding students in special education, however, has been challenged by school districts to ensure that students with disabilities receive the benefits of NCLB (Ferrell, 2005). Continuing the exemption of visually impaired students from taking standardized tests would negatively affect students with disabilities by sheltering them from the benefits of NCLB (Bolt & Thurlow, 2004; Ekstrom, 1998; Ferrell, 2005; West, 2005). West (2005) stated that professionals in the field of visual impairment should capitalize on the mandates of NCLB and allocate resources to improve educational materials, practices, assessments, and services. Although students with disabilities

differ in many respects from children without disabilities, Ekstrom (1998) stated that these differences should not interfere with information on policy, educational outcomes, access to the general education curriculum, or the availability of norm-referenced testing materials. Despite favorable arguments for including students with visual impairments in high-stakes testing, educators must consider how test scores impinge on decisions about individuals and institutions and the validity of the test results when making high-stakes decisions.

EFFECTS OF TESTING ON INDIVIDUALS AND INSTITUTIONS

Under NCLB, less than 1% of enrolled students may be tested using alternative assessments and removed from the calculation of adequate yearly progress (Browder et al., 2003). Yet, as McMahan (2005) pointed out, specialized schools often have a large percentage of children who are visually impaired and have additional disabilities who are given alternative assessments. McMahan stated, "Since only 2% of their enrollment can be excluded from the calculation of AYP, these schools are inappropriately labeled as under-performing schools, which could eventually result in a school closing" (p. 679). Other consequences of being labeled as underachieving include reduced funding and state intervention (NCLB, 2001).

At the individual level, the goals of NCLB are noble (Ferrell, 2005). Through NCLB, Congress attempted to create equality of opportunity and outcomes, regardless of socioeconomic status, ability, or cultural and linguistic differences (NCLB, 2001). For students with disabilities, NCLB has aimed to address the gaps in achievement between students in special education and students in regular education by holding schools accountable for the achievement of all students, as measured by standardized test scores (West, 2005). The consequences of low performance for individuals could include retention at each grade, no promotion from elementary school to middle school or from middle school to high school, and the failure to obtain a high school diploma.

Currently, graduation from high school is dependent on passing a high school exit examination. Students who do not pass this examination are not being awarded high school diplomas.

The practice of administering tests once a year as a measurement of students' progress and academic success has posed many problems. Sechrest (2005) warned against restricting causal inference to a single measurement and argued for critical multiplism, or the use of multiple measures (Cook, 1985, cited in Sechrest, 2005; see also Shadish, 1994). Similarly, Genshaft et al. (1980) stated that examiners must consider the "whole child" and include such factors as cultural differences, environmental conditions, and motivation when judging students' performance. They argued that a comprehensive evaluation could be obtained only if educators evaluate and integrate several sources of data.

DISCUSSIONS ABOUT VALIDITY

There are unique challenges to measuring the achievement of students who are blind or have low vision. Reid (1998) stated that the content validity of testing materials must be examined for cultural or visual bias. Pressley (2003) contended that any alteration to a test threatens external validity, which he defined as the conditions under which the student takes the test. Since accommodations that are needed for students with visual impairments require changes to the test, thereby affecting external and construct validity, Ekstrom (1998) stated that educators must evaluate whether the accommodations altered the construct or changed the intent of the test question. The transcription of test items into braille, for example, may change the construct of the test item. Reid argued that an immediate translation of a test item is too simplistic. She contended that before a test item can be adapted for students with visual impairments, educators must ask what is being measured and why is it being measured.

Bolt and Thurlow (2004) stated that since braille tests--especially math tests whose questions use figures, graphs, and tally systems--are more difficult for students to comprehend, the scores on

these tests are not representative of students' understanding of the constructs. Rather, tests that have been transcribed into braille must reflect an equal level of difficulty and synonymous constructs as the original test items. Similarly, large-print accommodations must be adapted with exact representations of the constructs.

Threats to internal validity include factors that are associated with incongruent populations. Standardized tests typically have normative data that are based on children without disabilities and, hence, may not be representative of the ability or skill level of a child who is visually impaired. Educators have argued that the validity of standardized test scores is affected by atypical development in cognitive, sensory, motor, and emotional development (Baker & Koenig, 1995; Swallow, 1981; Warren, 1984). Additional factors confounding test measures have included differences in the degree of vision loss and visual function, efficiency, and stamina (U.S. Department of Education, 2001). Baker and Koenig (1995) emphasized the need for normative data on children who are blind or visually impaired, and stated that comparing the test results of sighted students with those of students who are visually impaired is inappropriate. Bolt and Thurlow (2004) concluded that if tests are going to be used to make high-stakes decisions, then educators must carefully examine the validity of these tests when accommodations and modifications are being made.

Because of the emphasis on assessment results (IDEIA, 2004; NCLB, 2001), information on the validity of tests and testing practices is important. A qualitative research design with an interview format was chosen to allow the participants to respond openly to questions and to permit me to ask follow-up questions.

Methodology

PARTICIPANTS

School sites were selected from the online telephone list of the

Council for Schools for the Blind. Of the 48 schools that were listed, 2 were dropped because they did not have an academic focus and their students were not taking states' mandated assessments. Thirty-seven schools for students with visual impairments were contacted and participants from 27 schools, representing 27 states, were interviewed. A total of 36 interviews were conducted, with some schools having multiple participants. All the participants were interviewed separately, except for one group interview with 3 participants. The participants were recruited on a voluntary basis and agreed to allow their responses to be described in this article. The sample included 10 teachers, 18 administrators, 6 assessment specialists, and 2 participants who declined to identify their positions.

INSTRUMENTATION

Nine questions guided the interview (see [Box 1](#)), and follow-up questions were posed to clarify the responses or gather additional information. The validity of the instrument was tested with trial interviews, after which I modified the questions. The resulting questions are listed in Box 1. Question 9 focused on the occurrence of learning media assessments and functional vision assessments. Because the information was not pertinent to the purposes stated in this article, it is not reported in the findings.

PROCEDURE

A combination of telephone, e-mail, and in-person interviews was used. The interviews were not tape-recorded, but careful notes were taken. The reliability of the participants' answers was checked for verification following each interview by sending an e-mail message summarizing the conversation to the respective participant or by allowing the participant to view the summary sheet following the in-person interview. Some participants sent further e-mail messages, letters, and samples of instruments.

DATA ANALYSIS

The interview data were coded for common threads and organized into the following categories: (1) statements about state-mandated testing; (2) perceptions of the reliability, validity, and usefulness of various assessments; (3) the frequency of instruments; (4) reasons for choosing assessment tools; and (5) general concerns regarding the assessment process. Furthermore, data on the reliability, validity, and usefulness of assessments were quantified using a frequency count for each occurrence. Interobserver agreement (IOA) was determined by calculating the number of matched responses between two raters. Both raters had to agree on whether a response was given or not. If a response was given, both raters had to agree on all three categories of reliability, validity, and usefulness for IOA to be acceptable. Dividing the total number of matches by 36 interviews yielded an IOA between the two raters of 86.67%.

Findings

STATE-MANDATED TESTING

The first question was developed to gather information on state-mandated testing. The participants were asked to name the standardized assessment tool that was required by state policy. An analysis of the data showed that all 27 states differed in the state-required tests, indicating that each state had adopted different assessment instruments. Furthermore, each state had adopted various types of measurements for alternate assessments, including portfolios, computerized questionnaires, informal skills inventories, and descriptive and observational reports. Exemption criteria that qualified students for alternate assessments also differed among the states.

In addition, the participants were also asked if policy required students to take the state-mandated test at their grade level, that is, whether the grade or level of the test corresponded to the student's chronological age. In response to this question, 28 participants reported that the test was being administered at the

grade level, 1 reported that the test could be administered at a different grade level until the 2005-06 school year, and 7 were not sure of an answer or thought that the question was not applicable to the unique population of the school.

PERCEPTIONS OF RELIABILITY, VALIDITY, AND USEFULNESS

In response to questions about the reliability, validity, and usefulness of assessments, 26 participants gave their opinions for each of the questions, 10 responded to selected questions, and 10 did not answer the questions. Eleven participants believed that the tests were reliable measures, but 19 thought that the tests were not valid, and 16 thought that the data from tests were not useful. One participant commented that a test can have high reliability, but not be valid for its intended purpose. The general feeling of the participants was that testing was a mandatory part of the school year, but the relevance of the scores was not applied to everyday instruction. Additional data from observations, criterion-referenced tests, and multiple measures were used to determine students' current levels of performance on IEPs and to show adequate yearly progress. No differences were found among the four groups who were interviewed (administrators, teachers, assessment specialists, or job title not disclosed) (see [Table 1](#)).

FREQUENCY OF INSTRUMENTS

The participants revealed that the five most commonly chosen tests (frequency > 8) included the Brigance Comprehensive Inventory of Basic Skills, Revised (CIBS-R); classroom observations, checklists, and benchmarks; Basic Reading Inventory (BRI); Stanford Achievement Test (SAT); and portfolio assessments. In the 100 responses, 26 different instruments were named (see [Table 2](#)).

An interesting finding was that the participants did not name expanded core curriculum (ECC) assessments as often as academic assessments. This finding could be a reaction to the first

few questions pertaining to academic content. However, open-ended questions should have yielded responses about ECC areas because of their relevance to the IEP process. Four areas of the ECC were named (orientation and mobility, technology, social skills, and compensatory skills), and they were mentioned only eight times.

REASONS FOR CHOOSING ASSESSMENTS

The participants gave several reasons why they chose various assessment tools. Seven participants mentioned that they chose assessments to determine students' performance at a grade level or in a continuum of skills. They were interested in tests that measured academic achievement and thought that students' ability and progress on specific skills was useful information for IEP reports. Many of these participants reported using criterion-based measures, such as CIBS-R, BRI, SAT, and KeyMath. The second most common reason for choosing a specific instrument, stated by four participants, was to present an accurate "picture" of the students' performance. These participants reported using classroom observations, checklists, and curriculum-based assessments. Only four participants maintained that they used a variety of assessment tools in combination to evaluate students' performance. This was a small number of respondents who put the theory of critical multiplism into practice. (Critical multiplism, a term coined by Cook [1985], expanded by Shadish [1994], and promoted by Sechrest [2005], when applied to the assessment process, is to use multiple test measures to describe a child's academic potential.)

Three participants claimed that they chose instruments on the basis of whether the tests were available in the appropriate testing medium for the students. Other reasons included determining students' needs, showing annual progress, or transferring equivalent scores to public schools. Further remarks were that the chosen tests "work" for students with visual impairments and that the assessments were chosen by teachers, administrators, or

school psychologists. Two participants did not state their reasons for selecting the instruments they used.

The data revealed two common reasons for choosing the CIBS-R and the BRI. First, some participants thought that these tests assessed a continuum of skills, the results of which accurately portrayed if a student's ability within the continuum could be established. Others stated that these tests were chosen because of their availability in braille. Additional reasons for choosing the CIBS-R were that it had always been used at the school site and that the administrators recommended the test. The reported disadvantages were that the test yielded inflated scores and that the overuse of the instrument was leading to a carryover effect, in which a student's previous performance and memory carry over to each administration, resulting in inflated scores. One teacher commented that, in her school, the diagnostician was not trained to determine if a measure was influenced by a student's visual functioning or cognitive ability or the visual bias of a test item. Therefore, this teacher thought that the reports written by the diagnostician did not provide reliable, valid, or useful information.

GENERAL CONCERNS ABOUT THE ASSESSMENT PROCESS

Many of the general concerns about the assessment process involved the validity, reliability, and usefulness of the data. Eight participants thought that there is no appropriate test for students who score below a measurable level on standardized tests but do not qualify for the state's alternate assessment, that the state test does not address these students' skills, and that the test results are useless. Five participants were apprehensive about visual biases in the tests that could affect the validity of the instrument. They were concerned that when a test item was transcribed into braille, the original intention of the test item would not be retained and that the test could not assess congruent (visual as well as tactile) skills. Four participants thought that students were "overtested" and that the overuse of a test affected the reliability and validity

of scores because of a carryover factor that inflated the scores, and because testing reduces instructional time. An additional four participants stated that testing instruments do not measure small increments of growth and, hence, do not produce useful results if a child's report indicates the same level of performance year after year. Last, some participants reported that three schools were seeking a better assessment for math, reading, or general academic content at the elementary school level.

Discussion

RELIABILITY, VALIDITY, AND USEFULNESS OF SCORES

Many participants thought that most testing instruments provide invalid scores and that the data from state tests are not useful for IEPs or day-to-day instructional planning. As has been stated in the literature, and as was affirmed by the participants' answers, most standardized tests are not developed for students who are visually impaired; thus, differences are not factored into the normative data. A general consensus was that few tests have been developed to meet the needs of or to assess the skills of students who are visually impaired.

The literature has been consistent with the participants' concerns about the validity of test scores. Educators should choose instruments with these three points in mind: (1) educators need to ask what is being assessed and why it is being assessed, (2) educators need to be vigilant about oversimplifying accommodations that require braille transcription or large print, and (3) those who interpret the results need to be aware of the original construct of the test and how accommodations to the test may affect its validity.

The findings indicated that these cautions may not be considered when instruments are chosen. For example, if an assessment is being used because it was chosen by the school administrator, one must ask if the measurement is giving useful information for its intended purpose. If the measurement is not providing useful

information, perhaps an alternate one should be used. Similarly, choosing a test on the basis of its availability in braille may not be considered a best practice. A different solution may be to keep ongoing observational data, maintain students' portfolios, and use benchmarks from the general education standards.

The participants clearly indicated that the most useful data come from criterion-referenced tests (such as CIBS-R, BRI, SAT, and KeyMath). These tests provide useful information about how a student is performing within a continuum of skills, and teachers can use the information in annual IEP reports. One area of need that emerged from the responses is for a criterion-referenced test for reading or math with small increments that can show slower-than-average progress and that is appropriate for students who perform several grades below average. Educators face the challenge of choosing an appropriate criterion-referenced test when a limited number of such tests are produced in braille.

USE OF MULTIPLE MEASURES

Although researchers and the participants of this study have indicated that multiple measures are necessary (Genshaft et al., 1980; Sechrest, 2005), the policy trend is for a single measurement of achievement of students. The theoretical framework guiding policy decisions does not coincide with the theory used to assess students with visual impairments. Educators of students who are visually impaired understand individual differences and how these differences affect the development of cognitive, social, emotional, behavioral, and academic skills. They advocate for the use of critical multiplism. However, policy makers have enforced the use of a single measurement. For stakeholders, standardized testing is a momentary snapshot of a child's ability, and a student's achievement can be quickly compared to that of other students at the same grade level using normative data. When large groups are assessed, a single student's low performance will not upset the overall group score. However, for a particular low-scoring student, poor performance leads to

decisions about retention, promotion, and the awarding of a high school diploma. When such high-stakes decisions are being made on one measure, educators must advocate for equal access to testing media, access to the general education curricula, and fair accommodations for students with visual impairments.

EFFICIENCY OF TIME

Four participants stated that they were concerned that their students were being overtested. The literature and the participants' responses indicate that multiple measures for each student are necessary for a comprehensive description of achievement. Hence, the potential for overtesting is likely and consequential.

Overtesting can affect validity. Students are often tested with the same instruments each year. The validity of the measurements is affected if instruments are used repeatedly. Scores could show a carryover effect, which can be seen in IEPs if students have been tested using the same reading passages or math problems several times a year for several continuous years.

Furthermore, overtesting can have an unintended consequence that affects the amount of time during which educators are providing instruction. Educators must evaluate the consequences of testing on the amount of time for instruction and pose the question, How much time is spent testing versus teaching? They must decide whether the scores obtained from testing give valuable information that outweighs the time spent on testing that could be used for intervention and instruction. Another factor that must be considered is the size of caseloads. Because many standardized tests need to be administered individually, the amount of time spent in assessment must be a factor in determining the size of a caseload of a teacher of students with visual impairments.

LIMITATIONS

The study had three major limitations. First, not all the

participants answered all the questions because the interviews were sometimes cut short. Second, open-ended questions were used to eliminate researcher's bias, but the responses may have been affected by the participants' limited recall when responding. Third, the question on reliability, validity, and usefulness was not disaggregated into separate questions about each instrument. Therefore, the data are reflective of the assessment process in general, and conclusions about individual tests cannot be inferred.

The study was also limited by the nature of the survey data. The intent of the study was to discover general information about the assessment process. The participants stated their opinions--not facts--about the reliability, validity, and usefulness of instruments. Furthermore, the results of the group interview may not be as reliable as the individual responses.

Conclusion

Data from this study have implications under the mandates of current legislative policies (IDEIA, 2004; NCLB, 2001). NCLB threatens the individuality of IEPs by holding students accountable for achieving annual progress and grade-level success. Thus, there is a dichotomy between annual progress and individual progress. As special educators, we can no longer only assess an individual's skills. Instead, we must look at whether students with visual impairments are performing at grade level according to a norm-based set of standards and continuum of skills. Given the high expectations set forth in NCLB, educators of students with visual impairments must minimize the time spent on testing and maximize the time spent on instruction and intervention programs.

Specialized schools for students who are visually impaired also have an interest in reevaluating the available academic achievement tests and current standardization of test scores. The impact of performance results gathered from standardized academic achievement tests is considerable. The consequences of

not achieving adequate yearly progress include state intervention and the elimination of funding. However, decisions cannot be made about a school's performance on the basis of test scores that are invalidated for a multitude of reasons.

To be able to measure successfully the academic performance of a student who is visually impaired, new tests or measurements must be developed with normative data on students who are blind or have low vision. New tests must be designed with equal representation of the constructs represented in state standards, continuums of skills, and statewide standardized tests. Such normative data have not been compiled because the unique diversity of the field of visual impairment poses challenges for the construction of a normative sample. Perhaps the time has come for specialized schools for students who are visually impaired and educators of students who are visually impaired to converge and advocate for the development of an academic achievement test that is equivalent in esteem to the high expectations set forth in NCLB and state-mandated testing.

References

- Baker, C. P., & Koenig, A. J. (1995). Relationship of the Blind Learning Aptitude Test to braille reading skills. *Journal of Visual Impairment & Blindness*, 89, 440-448.
- Bolt, S. E., & Thurlow, M. L. (2004). Five of the most frequently allowed testing accommodations in state policy: Synthesis of research. *Remedial and Special Education*, 25, 141-152.
- Browder, D., Spooner, F., Ahlgrim-Delzell, L., Flowers, C., Algozzine, B., & Karvonen, M. (2003). A content analysis of the curricular philosophies reflected in states' alternate assessment performance indicators. *Research and Practice for Persons with Severe Disabilities*, 28, 165-181.
- Cook, T. D. (1985). Post-positive critical multiplism: Social science and social policy. In L. Shotland & M. Mark (Eds.),

Social science and social policy (pp. 21-62). Beverly Hills, CA: Sage.

Ekstrom, R. (1998). *Test interpretation and diversity: Achieving equity in assessment*. Paper presented at Assessment '98: Assessment for Change--Changes in Assessment, St. Petersburg, FL.

Ferrell, K. A. (2005). The effects of NCLB. *Journal of Visual Impairment & Blindness*, 99, 681-683.

Genshaft, J. L., Dare, N., & O'Malley, P. L. (1980). Assessing the visually impaired child: A school psychology view. *Journal of Visual Impairment & Blindness*, 74, 344-350.

Individuals with Disabilities Education Improvement Act (P.L. 108-446). (2004). Retrieved October 28, 2006, from http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_public_laws&docid=f:publ446.108

McMahon, E. (2005). Civil rights legislation. *Journal of Visual Impairment & Blindness*, 99, 678-680.

No Child Left Behind Act of 2001: Public law print of P.L. No. 107-110. (2001). Retrieved January 4, 2007, from <http://www.ed.gov/policy/elsec/leg/esea02/index.html>

Pressley, M. (2003). A few things reading educators should know about instructional experiments. *Reading Teacher*, 57(1). Retrieved January 4, 2007, from http://www.readingonline.org/articles/RT/9-03_Column

Reid, J. (1998). Assessing the literacy of adults who are visually impaired: Conceptual and measurement issues. *Journal of Visual Impairment & Blindness*, 92, 447-452.

Sechrest, L. (2005). Validity of measures is not a simple matter. *Health Research and Educational Trust*, 40, 1584-1604.

Shadish, W. (1994). Critical multiplism: A research strategy and its attendant tactics. *New Directions for Program Evaluation* 60, 13-57.

Swallow, R. M. (1981). Fifty assessment instruments commonly used with blind and partially seeing individuals. *Journal of Visual Impairment & Blindness*, 75(2), 65-72.


U.S. Department of Education. (2001). Policy guidance on educating blind and visually impaired students. *RE:view*, 33(2), 77-94.

Warren, D. H. (1984). *Blindness and early childhood development* (rev. ed.). New York: American Foundation for the Blind.

West, J. (2005). An opportunity slipping away? *Journal of Visual Impairment & Blindness*, 99, 677-678.

Cheryl Kamei Hannan, M.A., doctoral student, Department of Special Education, Rehabilitation, and Psychology, University of Arizona, and assistant professor, California State University-Los Angeles; mailing address: 5151 State University Drive, Los Angeles, CA 90032; e-mail: <cherylkamei@hotmail.com>.

• • • [Download braille-ready file](#)

 [Download ASCII text file](#) (ASCII files are for download only)

 [Download PDF file](#)

[Previous Article](#) | [Next Article](#) | [Table of Contents](#)

JVIB, Copyright © 2006 American Foundation for the Blind. All rights reserved.

[Search JVIB](#) | [JVIB Policies](#) | [Contact JVIB](#) | [Subscriptions](#) |
[JVIB Home](#)

If you would like to give us feedback, please contact us at
jvib@afb.net.

www.afb.org | [Change Colors and More](#) | [Contact Us](#) | [Site Map](#) |

Site Search

Go

[About AFB](#) | [Press Room](#) | [Bookstore](#) | [Donate](#) | [Policy Statement](#)

Please direct your comments and suggestions to afbinfo@afb.net
Copyright © 2006 American Foundation for the Blind. All rights reserved.