

# Large-Scale Academic Achievement Testing of Deaf and Hard-of-Hearing Students: Past, Present, and Future

Sen Qi<sup>\*1</sup>, Ross E. Mitchell<sup>2</sup>

<sup>1</sup>Gallaudet University

<sup>2</sup>University of Redlands

Received February 16, 2011; revisions received May 12, 2011; accepted May 16, 2011

The first large-scale, nationwide academic achievement testing program using Stanford Achievement Test (Stanford) for deaf and hard-of-hearing children in the United States started in 1969. Over the past three decades, the Stanford has served as a benchmark in the field of deaf education for assessing student academic achievement. However, the validity and reliability of using the Stanford for this special student population still require extensive scrutiny. Recent shifts in educational policy environment, which require that schools enable all children to achieve proficiency through accountability testing, warrants a close examination of the adequacy and relevance of the current large-scale testing of deaf and hard-of-hearing students. This study has three objectives: (a) it will summarize the historical data over the last three decades to indicate trends in academic achievement for this special population, (b) it will analyze the current federal laws and regulations related to educational testing and special education, thereby identifying gaps between policy and practice in the field, especially identifying the limitations of current testing programs in assessing what deaf and hard-of-hearing students know, and (c) it will offer some insights and suggestions for future testing programs for deaf and hard-of-hearing students.

We begin this review of deaf and hard-of-hearing student participation in large-scale assessments of academic achievement with an overview of how state and national testing ideology and policy in recent decades has intersected with testing research and practice. The developing dominance of test-based accountability has had a profound effect on the nature of student participation in testing programs and test use. The remainder of the paper is divided into three major sections chronologically: past performance, present

status, and the future of testing programs for deaf and hard-of-hearing students. Historically, we address what can be learned from the Stanford Achievement Test Series over the past three decades, both in terms of students' achievement and the test itself. For the present era, we examine the challenges faced in testing deaf and hard-of-hearing students in compliance with the new education laws, addressing especially such issues as whether the current testing programs are adequate and relevant. Looking to the future, we consider what testing systems should look like and what studies need to be done before such testing programs can be developed. This paper will advance the field toward a clearer strategy for the inclusion of deaf and hard-of-hearing students in state and national testing programs such that the test scores they obtain are reliable and provide valid inferences, which is all too often not the case presently.

## The Rise of Accountability

Our understanding of deaf and hard-of-hearing student participation and performance in large-scale academic achievement testing programs over the last few decades is informed by both a larger view of assessment policy history in American schools and particular developments unique to deaf education in the United States. The systematic inclusion and monitoring of deaf and hard-of-hearing students in large-scale assessment programs started in the 1960s, with the Metropolitan Achievement Tests and Stanford Achievement Test (SAT), before the current era of test-based accountability. Mazzeo (2001) refers to this

\*Correspondence should be sent to Gallaudet Research Institute, Gallaudet University, 800 Florida Avenue Northeast, Washington, DC 20002 (e-mail: sen.qi@gallaudet.edu).

earlier period as one when assessment was used primarily for *student guidance*. The purposes of testing were largely to diagnose student aptitudes and achievements, which would facilitate the identification and nurturance of gifted and talented students as well as detect problems requiring corrective action, and to inform appropriately differentiated instruction. National performance norms were essential for these identification and monitoring purposes. It was during this era that the Gallaudet Research Institute began its studies of the Stanford Achievement Test Series (hereafter referred to as the Stanford) for use with deaf and hard-of-hearing children.

As observers of public schooling in the United States may note, however, the expectations and uses of large-scale standardized tests were not constant over the last few decades. Although the Gallaudet Research Institute was developing strategies for improving the validity of scores derived from administration of the Stanford to deaf and hard-of-hearing students, state and later federal assessment policies were changing. Schools and students were becoming formally accountable for their performance on tests of academic achievement, first in the form of minimum competency tests and then to higher standards of “excellence” building upon momentum from the widely circulated *A Nation at Risk*, with its infamous decry of a “rising tide of mediocrity” (see, e.g., Bloomquist, 1986; Mazzeo, 2001). That is, just as the large-scale concerted effort to facilitate the participation of deaf and hard-of-hearing students in standardized tests of academic achievement within the student guidance framework of the mid-twentieth century had become established and recognized nationally, new uses and purposes of large-scale assessments were dawning with the era of test-based accountability.

Accountability for schools serving students with disabilities, including deaf and hard-of-hearing students, is now firmly entrenched in federal legislation that requires the inclusion of all students in state and district-wide assessment programs (i.e., the No Child Left Behind Act of 2001, hereafter NCLB, and Individuals with Disabilities Education Improvement Act of 2004, hereafter IDEA '04). According to these laws, large-scale academic assessments are to “measure the academic achievement of such students relative to

State academic content and State student academic achievement standards” (NCLB, Title I, Part A, Subpart 1.b.3.C.ix.II, 2002), and any special assessment needs of students with disabilities are to be met “with appropriate accommodations and alternate assessments where necessary and as indicated in their respective individualized education programs [IEPs]” (IDEA '04, Title I, Part B, Section 612.a.16.A, 2004). Moreover, test quality standards have been imposed. These academic assessments are to “be used for purposes for which such assessments are valid and reliable and be consistent with relevant, nationally recognized professional and technical standards” (NCLB, Title I, Part A, Subpart 1.b.3.C.iii, 2002). These new laws and regulations call for high-quality tests or adequate test accommodations to meet the need of students with disabilities, including deaf and hard-of-hearing students.

At the same time, however, special education law recognizes that technical standards have not been established for the inclusion of special populations in academic assessment programs, which threatens the legitimacy of test-based accountability for all students. This is largely a consequence of the need to provide testing accommodations for students with disabilities in order for them to participate in large-scale assessment programs. Sireci, Scarpati, and Li (2005) described how a number of popular test accommodations have little or no consistent evidence to justify their use. This is why states are authorized to use IDEA '04 funds “to support the development and provision of appropriate accommodations for children with disabilities, or the development and provision of alternate assessments that are valid and reliable for assessing the performance of children with disabilities, in accordance with [NCLB]” (IDEA '04, Title I, Part B, Section 611.e.2.C.x, 2004). Further, the recently established National Center for Special Education Research is responsible for promoting the improvement of “the alignment, compatibility, and development of valid and reliable assessments, including alternate assessments, as required by [NCLB]” (IDEA '04, Title II, Part E, Section 177.a.5, 2004; also see Title I, Part B, Section 663.b.2, 2004). These federal mandates for test use, research, and development, as well as the historical changes in the primary uses and purposes of testing in schools, serve to frame our

discussion of deaf and hard-of-hearing students' participation in and performance on large-scale assessments of academic achievement.

#### Assessment Challenges for Deaf and Hard-of-Hearing Students

Developing valid and reliable assessment instruments to measure the achievement levels of deaf and hard-of-hearing students is an ongoing project. Allen, White, and Karchmer (1983) alerted the field to problems with using standardized tests designed for hearing students when assessing deaf and hard-of-hearing students over 20 years ago. Two important issues stand out. First, a significant proportion of deaf and hard-of-hearing students are neither on grade level nor receiving undiluted instruction in the general curriculum (also see Mitchell, 2008; Steffan, 2008). Holt and Allen (1989) noted that deaf and hard-of-hearing students who attended special programs or schools may have had a curriculum that varied significantly from the general curriculum, which is likely to have remained true for at least some time after the general curriculum was mandated for students with disabilities in the *Individuals with Disabilities Education Act Amendments of 1997* (hereafter referred to as IDEA '97). As a result, deaf and hard-of-hearing students' opportunity to learn the content represented on grade-level tests was and may continue to be unequal to that of the general population. The assessment consequence of this differential opportunity to learn for deaf and hard-of-hearing students is that when given standardized tests for their age or grade, these students would obtain very low and unreliable scores. Out-of-level testing procedures were developed to reduce the threat to test validity resulting from special curricula for deaf and hard-of-hearing students (Allen, 1984).

Second, many deaf and hard-of-hearing students receive classroom instruction through sign language (American Sign Language, or ASL, in the United States), or in some other visual communication mode (e.g., Karchmer & Mitchell, 2003; Mitchell, 2004; Mitchell & Karchmer, 2005). For these students, English is not their primary language of academic discourse. Many of these students who sign have limited English proficiency as well (e.g., Bosso, 2008; Emmorey,

Bellugi, Frederici, & Horn, 1995; Johnson, Liddell, & Erting, 1989; Jones, 2008; Kelly & Barac-Cikoja, 2007; Mayberry & Lock, 1998; Meadow-Orlans, Mertens, & Sass-Lehrer, 2003; Morford & Mayberry, 2000). Therefore, testing in the English language may inhibit the ability of deaf and hard-of-hearing students to fully express what they know. However, the potential solution of translating standardized tests into ASL has not been widely adopted for large-scale testing due to lack of psychometric studies (Allen & Sligar, 1994), as well as practical considerations, such as funding. The South Carolina Department of Education has overcome the practical concerns by using department funds to contract for the development of videotaped ASL and Signed English translations of state achievement tests for use with deaf students (Foster, 2008; South Carolina Department of Education, 2005), though documentation of the translated tests' validity and reliability remains unavailable. Systematically developed and standardized ASL presentations of state and district-wide assessments remain unavailable from test developers and vendors and have been considered for development in only a few other jurisdictions throughout the United States (for exceptions, see Bello, Costello, & Recane, 2008; Maihoff et al., 2000; Tindal, 2006).

#### Past Performance

Despite the challenges and difficulties, deaf and hard-of-hearing students in the United States have been assessed using standardized tests, most frequently the Stanford, for over three decades (for dates and samples sizes, see Table 1). Beginning in 1969, about 12,000 deaf and hard-of-hearing students were tested using the fifth edition of the Stanford. In this first attempt at large-scale assessment research with this population, the Stanford was adopted without any

**Table 1** Dates and sample sizes of past special norming studies for the Stanford test

Test edition	Edition year	Norming year	Norming sample size
6th	1973	1974	6,873
7th	1982	1983	8,311
8th	1989	1990	6,932
9th	1996	1996	4,810
10th	2003	2003	3,569

modification or revision. However, the researchers were fully aware of the limitations of such an approach. Item analyses were conducted to gather information for future norming studies (Office of Demographic Studies, 1969).

In 1974, a special version of the Stanford was developed for deaf and hard-of-hearing students (Stanford Achievement Test for Hearing Impaired [SAT-HI]) through a special norming procedure (Trybus and Karchmer, 1977). The SAT-HI was still based on the Stanford for hearing students, but it made significant progress in improving the validity of achievement testing for deaf and hard-of-hearing students. The SAT-HI not only used a nationally representative sample of its target population for its norming study but also took the following measures to standardize the testing procedures to address special issues for deaf and hard-of-hearing students: A screening test was adopted to ensure each student was tested at the appropriate grade level, by subject, to minimize guessing; based on the screening tests, mixed (out of level) test batteries were assigned (e.g., a 14-year-old student could take the third-grade reading testing test and the sixth-grade mathematics test); practice test materials were provided to familiarize deaf and hard-of-hearing students with the test's format and how to record responses (this procedure was later adopted by the test publishers for the general population as well); and the dictated test instructions normally read aloud to young hearing students were printed on the test booklets. Test administrators directed the deaf and hard-of-hearing students to read the printed instructions before answering questions, making sure the examinees understood what they were expected to do. Such accommodation attempted to provide an equal opportunity for the deaf and hard-of-hearing students to understand tasks on the test as would their hearing counterparts, without altering the contents and presentation of the test questions. The aforementioned special testing procedures continued to be adopted in the subsequent editions of the Stanford tests for deaf and hard-of-hearing students.

Since the SAT-HI, special norming studies have been conducted for all four subsequent editions of the Stanford (Allen, 1986a; Holt, Traxler, & Allen, 1992, 1997; Mitchell, Qi, & Traxler, 2007). Together, these

five norming studies represent a rich source of historical data over the last three decades. The results of the norming studies depict a picture of the past and present status of deaf and hard-of-hearing students' academic achievement. It should be noted that the norming sample sizes for the last two test editions were smaller. It was due to the higher nonparticipation rates as a result of more testing burdens in the schools. The smaller samples, however, still provided good representations of the population, although the precision of the estimates may be reduced.

### Review of the Stanford Results

Because the Stanford has been the only national large-scale assessment regularly used to monitor the academic achievement of deaf and hard-of-hearing students over the last three decades, results from this test are used to examine the achievement record leading up to the present era. The historical data used in this report are from archives of the Gallaudet Research Institute, which conducted the special norming studies of the Stanford. Data from all five previous norming studies are reviewed. Table 1 presents the dates and sample sizes for the norming studies.

The results from the norming studies have been published (Gallaudet Research Institute, 1983, 1991, 1996, 2004; Jensema, Schildroth, & O'Rourke, 1974) or documented in an internal report (Mitchell et al., 2007). Age-based percentile ranks for each subtest, as well as conversion tables for raw scores, scaled scores, and grade-equivalent scores, serve as the data for this study.

*Data selection and computation.* The subtests contained in the Stanford batteries vary between test editions and test levels. For example, the 10th edition of SAT has 11 test levels for students from 1st grade to 11th grade (Primary 1 for grades 1 and 2; Primary 2 for grades 2 and 3; Primary 3 for grades 3 and 4; Intermediate 1 for grades 4 and 5; Intermediate 2 for grades 5 and 6; Intermediate 3 for grades 6 and 7; Advanced 1 for grades 7 and 8; Advanced 2 for grades 8 and 9; Test of Academic Skills 1 for Grade 9; TASK 2 for Grade 10; and TASK 3 for Grade 11).

To facilitate comparisons, this study reviews only the reading comprehension, mathematics problem-solving, and mathematics procedures' subtests. Almost all Stanford editions and test levels contain these three subtests. An important exception is the inclusion of the three TASK level tests in the 2003 study of the tenth edition. The TASK level tests offer only a single total mathematics scale and not the two subtests administered for the Primary 1 through Advanced 2 levels. For each student who took the TASK level tests, mathematics problem-solving and mathematics procedures' subscores were imputed by regression equations, in which mathematics total score, age, and testing level were used to predict mathematics subtest scores.

The distribution of academic achievement among deaf and hard-of-hearing students within each age or grade deviates from that of their hearing peers because the students in the norming sample were often assigned a test below their age or grade level. This out-of-level testing was necessary because some deaf and hard-of-hearing students lagged behind their hearing counterparts, and their development was uneven across different subject areas. If they were given higher level tests, too far beyond what their actual abilities could handle, both validity and reliability of their scores would be in question. Such students' performances are better summarized by age instead of grade. The median (50th percentile) scale scores of each age group (8–18 years) were obtained for the three subtests. These median scaled scores were then converted to grade-equivalent scores using tables provided by the test publisher.

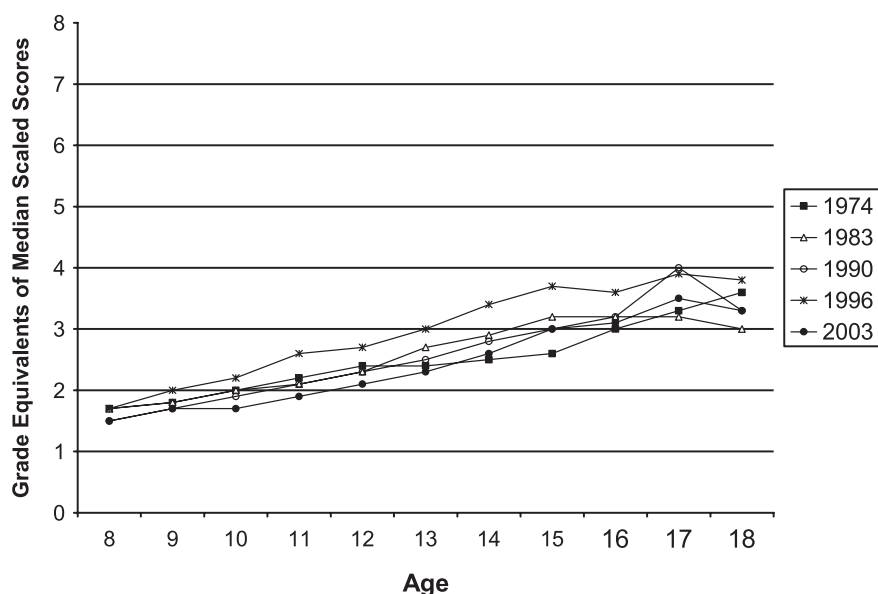
It should be noted that the Level 1 and Level 2 tests in the 1974 battery, as well as the Primary 1 and Primary 2 tests in the 1983 battery, gave only total reading scores. So the median reading comprehension scaled scores for lower age groups were found from the norming tables for Level 3 or Primary 3 tests. Because the scaled scores have equal units on a continuum that covers the full test range, this substitution should be adequate. Similar to the reading comprehension tests, the Level 1 test in the 1974 battery, as well as the Primary 1 test in the 1983 battery, did not give two separate subscores for mathematics. The scaled scores for the two mathematics

subscales were obtained from the norm tables for higher level tests. It should also be mentioned that before the 1996 edition, the two mathematics subtests were labeled "mathematics application" and "mathematics computation." In the 1996 and 2003 editions, these two subtests were renamed "mathematics problem solving" and "mathematics procedures." The two sets of terms represent similar constructs in mathematics learning. For simplicity of presentation, the latter terms are used here.

*Results.* The compiled data are presented in three trend figures. For all three figures, the grade equivalents of median scaled scores are plotted against student age. By design, the relationship between age and grade equivalent for hearing students are set to be constant. For example, the median scale scores of 8-year-old hearing students would have a grade equivalent of 3, the typical grade of enrollment for children of that age. Therefore, the lines for hearing students were not plotted in the figures because they would have been identical for all years and all subtests. For comparison purposes, when reading these figures, one may assume that the hearing students from 8 to 17 years of age have corresponding grade equivalents from 3 to 12, respectively.

Figure 1 shows the results for student performance on reading comprehension tests over three decades. The narrow band formed by the five lines in this figure indicates that the normative performance of deaf and hard-of-hearing students on reading comprehension tests has been remarkably consistent. Their performance levels are slightly higher for each age cohort from age 8 through age 17, but median performance never exceeds the fourth-grade equivalent for any cohort.

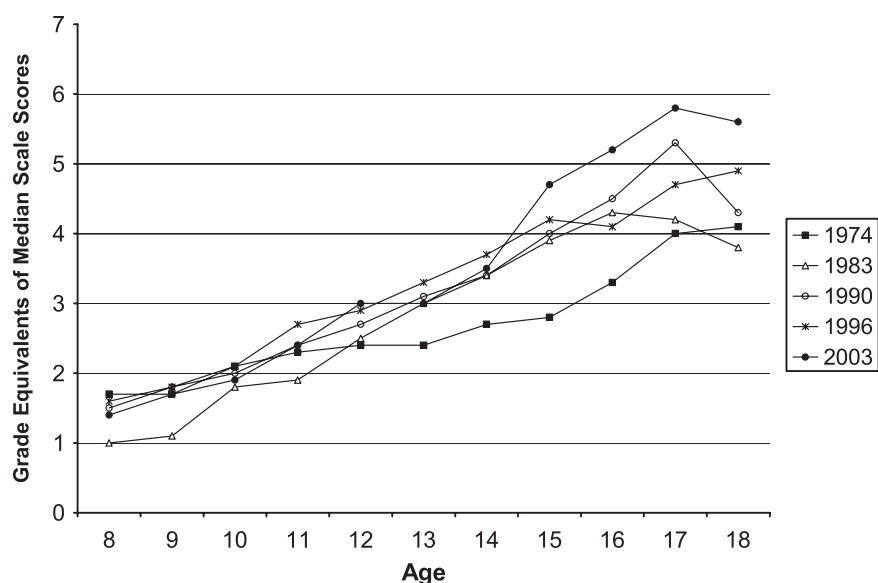
Figures 2 and 3 show the performance for deaf and hard-of-hearing students on mathematics problem-solving and mathematics procedures subtests, respectively, over the same three decades. In mathematics problem solving, unlike in reading comprehension, older students achieve notably higher levels of performance, especially with more recent test editions. Median performance for 17-year-olds approaches the equivalent of sixth grade, two grade equivalents higher than for reading comprehension.



**Figure 1** Grade equivalents of median scaled scores on Stanford Achievement Test for Deaf and hard-of-hearing student norming samples in the United States, by age, 1974–2003: reading comprehension.

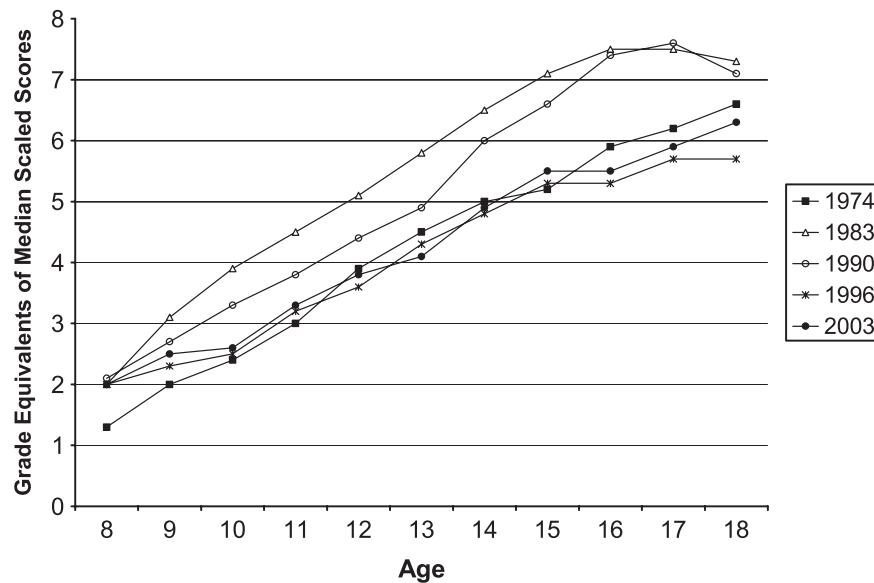
The performance trend in mathematics procedures is very consistent for the 6th, 9th, and 10th editions (1974, 1996, and 2003), as shown in Figure 3. However, the performance profiles for the 1983 and 1990 norming groups are higher than the others. By age 16 and 17, the medians of the 1983 and 1990 samples reached the grade equivalent of 7.5, whereas

the norming groups in other years only achieved about the sixth-grade level. Allen (1986b) analyzed the differences between 1974 and 1983 norming procedures and confirmed a plausible gain in student performance within that decade. However, there has been no comparative analysis after 1986. For the performance regression after 1990, sampling and norming procedural



**Figure 2** Grade equivalents of median scaled scores on Stanford Achievement Test for Deaf and hard-of-hearing student norming samples in the United States, by age, 1974–2003: mathematical problem solving.





**Figure 3** Grade equivalents of median scaled scores on Stanford Achievement Test for Deaf and hard-of-hearing student norming samples in the United States, by age, 1974–2003: mathematical procedures.

differences are among the possible explanations. But before any conclusion is made, the specific contents and features of the 1983 and 1990 test editions must also be closely examined.

Significance tests were not conducted between each pair of medians in these figures. The different editions of the tests were given at disparate times, to different student norming groups, with different testing procedures (screening tests, language accommodations, time and setting accommodations, etc.). Direct comparison between each pair of tests is not sensible, especially because there was no linking or equating between all pairs of tests. However, as a figurative rather than legitimately statistical analysis, the standard errors of measurement of those tests are approximately one grade equivalent, so some cohort means on the mathematics subtests would exceed the 95% confidence interval of another. Differences of roughly two grade equivalents can be seen between the 1974 and 2003 editions of the mathematics problem-solving subtests (Figure 2) and the 1983 and 1996 editions of the mathematics procedures subtests (Figure 3) for 15- and 16-year-old students.

In sum, according to the SAT, three patterns emerge from the historical trends in reading and mathematics achievement. First, the performance of deaf and hard-of-hearing students has been consis-

tently below hearing students. There is an achievement gap. Second, the gap is larger for reading than for mathematics. And third, the gaps between deaf and hard-of-hearing students and hearing students have not been closing over the last three decades, with the possible exception of mathematics problem solving. As measured by the well-known and widely used Stanford Achievement Test Series, there has been little or no change in the central tendency of academic achievement among the deaf and hearing student population over the last three decades.

### The Present Status

Before IDEA '97 and NCLB, the out-of-level testing procedures developed by the Gallaudet Research Institute provided at least a partial solution to the problem of valid measurement by employing a screening test, in conjunction with teacher judgment, to identify the grade-level test most appropriate for the student's instructional and performance levels. However, out-of-level testing is no longer generally acceptable because an out-of-level test is, by definition, not the regular assessment for students who are striving to "achieve passing marks and advance from grade to grade" in conjunction with state standards (see Rehnquist opinion for the court, 1982, in 458

US 176, *Board of Education of the Hendrick Hudson Central School District v. Amy Rowley*). As stated in a notice of proposed rule making (NPRM):

The current [NCLB] Title I regulations do not prohibit the use of out-of-level assessments in all cases. They may be used to assess students with the most significant cognitive disabilities if they are aligned with a State's alternate achievement standards that meet the requirements of current [regulations]. (U.S. Department of Education, 2005, p. 74627)

The field of deaf education is now confronted with a professional dilemma by employing out-of-level testing. Though it is consistent with the student guidance perspective for schools to use the Stanford and its screening tests for out-of-level test assignment as a supplemental assessment to monitor the academic achievement of deaf and hard-of-hearing students in accordance with their IEPs, the practice is now stigmatized as an assessment strategy only appropriate for students with severe cognitive disabilities (e.g., Cawthon & Wurtz, 2009; Jones, 2008; Moore, 2008; Steffan, 2008). Clearly, out-of-level testing has no comfortable place in the current test-based accountability regime. Nonetheless, it may still be a useful response to the practical assessment needs of educators responsible for the deaf and hard-of-hearing students in special education who are not keeping up with the normative grade progression through the general curriculum.

Recent assessment evidence from the national evaluation that followed the implementation of IDEA '97, for which data were collected during the first years of the current century, provides confirmation of the achievement patterns observed in the three decades of Stanford results (Mitchell, 2008). For example, based on a summary of the Special Education Elementary Longitudinal Study findings (Blackorby and Knokey, 2006), which were based on administration of the Woodcock-Johnson III with IEP-specified accommodations, students with an identified (or documented) hearing loss scored significantly lower than their hearing peers on tests of reading passage comprehension and mathematics computation. In addition, similar to that shown above for the Stanford, the achievement

gap in reading is much larger than that observed for mathematics. Though the IDEA '97 evaluation studies and the Gallaudet Research Institute studies used different standardized assessments to measure academic achievement, sampling frames for recruiting participants, and directions for how to administer the assessments, their target populations were the same and, more importantly, their assessment results are quite similar as well.

More recent data from state-level assessment programs confirm the persistent trend of low academic achievement of deaf and hard-of-hearing students. In California, only 8% of deaf students and 15% of hard-of-hearing students scored proficient or advanced on the California Standards Test for English Language Arts, and for Mathematics, 10% of deaf students and 18% of hard-of-hearing students scored proficient or advanced (California Department of Education, 2007). We have not identified any other statewide summaries in the literature, but there are several school-level reports from state schools for deaf. Cawthon (2008) summarized the student proficiency rates from schools for the deaf in 21 states that had report cards for deaf students in 2007. For example, only 15.6% of students in the Louisiana School for the Deaf (LSD), across all grades, achieved proficiency in reading and 31.3% achieved proficiency in mathematics (Cawthon, 2008, p. 106). To put these statistics into context, compare statewide and LSD eighth-grade reading and mathematics proficiency levels in 2007, for example, statewide they were 58.9% and 55.9%, respectively, whereas at LSD they were 23.8% and 19.0%, respectively (CCSSO SchoolMatters, 2007). School and statewide comparisons can be made for less commonly tested subjects as well: 20% of LSD students taking the state's Graduate Exit Examination (GEE) achieved at or above the basic level on the science test, compared with 60% statewide, whereas only 10% achieved at or above basic level on the GEE social studies test, compared with 64% statewide (SchoolDigger.com, 2010). Though the magnitude of the achievement gap between students at schools for the deaf and students statewide varies from state to state, the differences are always substantial and, with few exceptions, deaf and hard-of-hearing



students' proficiency rates are quite low (e.g., see Cawthon, 2008; <http://www.SchoolMatters.com>).

In the present era of accountability, deaf and hard-of-hearing students who do not have severe cognitive disabilities are included in their state's assessment programs largely by accommodation. Clapper, Morse, Lazarus, Thompson, and Thurlow (2005) surveyed state policies and found the most common accommodations for deaf and hard-of-hearing students are ASL interpreter, extended time, separate location, and computer administration. As identified by Cawthon (2006, 2007) and Cawthon and Online Research Laboratory (2008) in more recent surveys, the most popular accommodations in use with deaf and hard-of-hearing students are extended time, followed by an interpreter for test directions, and a separate testing room. Less popular and controversial, ASL presentation of the test items themselves is also utilized as an accommodation; however, some states forbid ASL presentation of the entire statewide assessment, especially the part testing reading comprehension (Case, 2008; Lazarus, Thurlow, Lail, Eisenbraun, & Kato, 2006). When a reason is given, ASL presentation is considered invalid either for the unfortunate lack of psychometric evidence supporting the comparability of scores from tests translated into ASL or because an ASL presentation is considered to fundamentally alter the intended construct to be measured (e.g., there is no reading comprehension of written English measured when everything is signed and nothing is read; also see Case, 2008). However, Cawthon (2007) reports that the three most recommended forms of test accommodation, based on teachers' opinions of best practice, are student signs to scribe, interpreter for test items, and read aloud. Educators in the field recognize the importance of the language accommodation. Nonetheless, validation studies in this area are still lacking.

Given the special linguistic considerations confronted in the education of these deaf and hard-of-hearing children who use of ASL for daily classroom discourse and instruction, it is important to note that both the general and special education laws acknowledge that assessments must accommodate America's multilingual school population. As set forth in NCLB:

Each State plan shall identify the languages other than English that are present in the participating student population and indicate the languages for which yearly student academic assessments are not available and are needed. The State shall make every effort to develop such assessments and may request assistance from the Secretary [of Education] if linguistically accessible academic assessment measures are needed. Upon request, the Secretary shall assist with the identification of appropriate academic assessment measures in the needed languages, but shall not mandate a specific academic assessment or mode of instruction (Title I, Part A, Subpart 1.b.6, 2002).

As set forth in IDEA '04:

Testing and evaluation materials and procedures utilized for the purposes of evaluation and placement of children with disabilities for services ... shall be provided and administered in the child's native language or mode of communication (Part B, Section 612.a.6.B). [Specifically,] assessments and other evaluation materials used to assess a child [shall be] provided and administered in the language and form most likely to yield accurate information on what the child knows and can do academically, developmentally, and functionally (Part B, Section 614.b.3.A.ii, 2004).

Though a deaf student who communicates primarily through ASL is not typically classified as a limited English proficient student, such classification would permit, under NCLB, "the local educational agency [to] make a determination to assess such student in the appropriate language other than English for a period that does not exceed two additional consecutive years [beyond the first three years enrolled in school], provided that such student has not yet reached a level of English language proficiency sufficient to yield valid and reliable information on what such student knows and can do on tests (written in English) of reading or language arts" (Title I, Part A, Subpart 1.b.3.C.x, 2002). In other words, federal law has provisions for the special assessment needs of deaf and hard-of-hearing students who sign that could reasonably be construed

to relieve states of the necessity to test these students using written English instruments, even for reading and language arts during the first few years of schooling. However, as noted previously, only South Carolina has systematically undertaken to provide statewide assessments in ASL (Foster, 2008). The present era remains dominated by the idiosyncrasies of locally translated administrations, which vary in quality and potentially invalidate test scores because of their lack of consistency from one administration to the next, if ASL translation or interpretation accommodations are utilized at all.

### The Future of Testing

If deaf and hard-of-hearing students are to be full participants in state and district-wide assessment programs then the tests used must be valid and reliable for them as well. As mentioned in the previous sections, two factors may contribute to the proficiency gaps between deaf and hard-of-hearing students and their hearing peers. The first is the opportunity to learn: The services provided to this group of students with special needs may not be adequate and effective, so they lag behind their hearing peers. The second is the assessment system: the tests being used to monitor achievement may be biased against this special population. Each factor may have its independent effect, but most likely the gaps result from a combination of the two. Solving issues about the opportunity to learn is largely beyond the responsibilities of testing professionals. In this section, we will focus on the second factor. We offer some suggestions for the future development of large-scale tests for deaf and hard-of-hearing students.

### Curriculum–Test Alignment

In developing large-scale assessments of academic achievement, one can never emphasize too much the importance of curriculum–test alignment for making valid inferences from test results. For a test to be truly educational, it must assess what has been taught in the classroom. The common expectation is that students' test scores reflect how well they have learned their lessons. For this interpretation to be true, the content of the test must match the content of the curriculum.

Moreover, the kinds of questions that are asked and the format in which they are presented must be familiar to students as a result of their instructional experiences. Otherwise, the strength of the relationship between schooling and test scores is diminished. Although test developers cannot guarantee instructional behavior in the classroom, they can strive to accurately represent the curriculum presented, including the common tasks students must perform. At the same time, educators must take responsibility for delivering the assessed curriculum, at a minimum, so that students have an opportunity to demonstrate that they know what is expected of them (as represented on the test). As long as there is a general assessment for all students regardless of whether they have an IEP, there is a reciprocal relationship between test developers, who must accurately represent the general curriculum, and classroom teachers, who must instruct students in the general curriculum.

The development of standardized large-scale assessments of academic achievement, ideally, would begin with a survey process: identify locally adopted curricula and related standards to understand what is being taught in the schools, observe instructional behaviors to determine how the curricula are experienced in the classroom, and attend to any accommodation practices or other variances from the modal program of instruction. Curriculum–test alignment greatly depends on the results of this survey process. The greater is the variety of responses and observations, the poorer is the curriculum–test alignment for any particular school or classroom. Test developers may be confronted with the uncomfortable reality that a single assessment instrument will not provide measurements that accurately reflect the teaching and learning happening in schools. In other words, in the absence of a uniform curriculum across classrooms and schools, which has been a serious issue for schools and programs serving deaf and hard-of-hearing students in the past, large-scale assessment becomes a measure of performance on the test itself and provides limited capacity to draw inferences about what has been taught and learned in schools. Because test-based accountability is likely to stay in the school systems, improving the performance of deaf and hard-of-hearing students on large-scale assessments of

academic achievement will depend on their opportunity to receive instruction in the general curriculum represented on the test.

The exception that shifts the greater burden back to test developers is alternate assessments. Though supposedly aligned in some fashion with regular achievement standards, alternate assessments for modified achievement standards (see NPRM noted above, U.S. Department of Education, 2005) or alternate achievement standards test curricula at variance with the general curriculum. Important research needs to be done in the area of alternate assessment. However, because alternate assessment for alternate achievement standards is for a small fraction of the student population with severe cognitive disabilities (up to 1%) and alternate assessment for modified achievement standards is for a larger, but still small fraction of the total population (up to 2%; all of whom have been identified for special education), we will not elaborate on this point further but continue to address assessment for regular achievement standards.

### Test Accommodation

Accommodation in the classroom has important consequences for test development. If all students, including those who are deaf and hard of hearing, participate in the general education curriculum, there will certainly be occasions when the modal program of instruction fails to make the tasks accessible to all students. Whether viewed as a matter of rights, obligations, or pragmatics, accommodations are necessary. Because the same or similar tasks are to be performed on academic assessments, test accommodations are necessary as well.

We must stress that the purpose of test accommodation is not to increase the scores for students who receive instructional accommodations in the classroom or who are otherwise identified for special education or related services. Test accommodation serves a fundamental role in establishing the validity of test score interpretations for students with disabilities. A test accommodation removes obstacles to successful test performance that are unrelated to the traits or constructs being measured. An appropriate test accommodation is intended to make the assessment equally

accessible for all students regardless of disability so that they all have an equal opportunity to fully demonstrate what they have achieved academically. In psychometric terms, test accommodation produces an “interaction effect” (Sireci et al., 2005). That is, the change in testing conditions makes a difference only to the students who would otherwise have their performance artificially depressed by construct irrelevant barriers. Students who experience no such barriers would perform equally well with or without the accommodation.

It is also necessary to note that not all test accommodations are equal or equally appropriate. A recent review highlights some of the controversies regarding the appropriateness of different accommodations available to students with disabilities (Sireci et al., 2005). For example, extended testing time is the most commonly used accommodation, but it may not be suitable for all situations. If a student does not understand the test instructions due to an insurmountable language barrier, adding time alone is not going to help. Also, if a test measures speed as an indicator for mastery of the skills in question, then extended time will give the accommodated students an unfair advantage. Test content is also an important consideration when selecting accommodations. For example, removing any English language barrier to demonstrating proficiency by translating the test into ASL would, in principle, be appropriate for certain types of mathematics questions, but not for a test of English reading comprehension. In the latter case, written English is a central feature of the construct being measured by the instrument. However, reading comprehension test items that require phonetic cues are highly unlikely to contribute to valid inferences about what deaf and hard-of-hearing students know. Also when mathematics problems are signed to the deaf and hard-of-hearing students, attention must be paid to whether the difficulty of items is changed as a consequence of the iconicity of signed languages (see Ansell & Pagliaro, 2001, 2006; Thurlow, Johnstone, Thompson, & Case, 2008). The key to adopting an appropriate accommodation for special populations is to carefully examine the barriers students may encounter during testing. By identifying and removing obstacles to task performance, test accommodations allow the students with disabilities to

fully demonstrate what they know and what they are able to do.

#### A Call for Research on Language-Based Accommodations

Understanding and development of test accommodations would greatly profit from a systematic program of research. We identify two lines of research, modified English and ASL adaptation, both essential to improving the measurement of academic achievement among deaf and hard-of-hearing students. Such accommodations depend on considering these students as having limited English proficiency. That is, deaf and hard-of-hearing students may be English language learners (ELLs) with prior proficiency in ASL (roughly 4% of this population is born to deaf parents who sign, as well as some additional but unknown percentage whose hearing parents have or develop signing proficiency with their children before they enter school; see Mitchell & Karchmer, 2005, Table 3; also Mitchell, 2004; Mitchell & Karchmer, 2004a, 2004b), or late language learners who have experienced delays in learning any language (Emmorey et al., 1995; Mayberry & Lock, 1998; Morford & Mayberry, 2000). Regardless of when and which languages they have acquired, limited access to the spoken English language environment around them adversely affects deaf and hard-of-hearing students' opportunity to learn the language from many common auditory channels such as radio broadcasts, spoken conversations, and so on, which are easily accessible to hearing students. Therefore, their lower performance on English tests could be an indication of such inadequate exposure to the language.

It should be noted that the analogy of hearing ELL may be imperfect for deaf and hard-of-hearing students. Unlike hearing students who learn a second spoken language, when deaf and hard-of-hearing students learn English, they face more challenges due to their delay in acquiring a first language (Marschark, Lang, & Albertini, 2002; National Research Council, 2005) and their lack of phonological knowledge in language comprehension (Kelly & Barac-Cikoja, 2007). Also, due to diverse communication modalities, the English proficiency of deaf and hard-of-hearing students varies widely (Case, 2008). Nonetheless,

approaching deaf and hard-of-hearing students' assessment needs as bearing greater similarity to those of ELLs than to the generic "student with a disability" could enhance the appropriateness and adequacy of accommodations for them, allowing deaf and hard-of-hearing students to demonstrate the knowledge and skills being tested independent of their English language ability.

*Modified English.* The first line of inquiry we propose is to explore the potential benefit from accommodations designed for hearing ELL's, particularly the modified (or simplified) English text strategy developed by Abedi and coworkers (e.g., see Abedi, 2005; Abedi, Hofstetter, & Lord, 2004). Although this accommodation has not been widely adopted in large-scale testing programs, the study results show that most ELLs preferred the modified English version of the tests over the original, and those who were tested in modified English performed better than those who were tested in the original version. Most significantly, unlike the extended time accommodation, the linguistic modification of test questions with excessive language demands proves to be the only accommodation that narrows the gap between ELL and non-ELL students (Abedi, Lord, Hofstetter, & Baker, 2000). Only one limited study, by Mowl (1985), has applied a modified English accommodation with deaf and hard-of-hearing students; it offers a suggestion of some promise for this approach. The Commonwealth of Virginia permits this accommodation for deaf and hard-of-hearing students on the statewide accountability assessment (Parent Resource Center Fairfax County Public Schools, 2004), but there have been no psychometric studies of its implementation.

*ASL adaptation.* Ostensibly, translating test questions and directions into an examinee's first language would seem to be the most sensible solution to the problems that ELLs have in testing. However, this is not necessarily the case. For example, Abedi, Lord, and Hofstetter (1998) found that when mathematics items were presented in Spanish to Spanish-speaking students, those who received instruction in English performed significantly lower than those who received instruction in Spanish. In other words, it may only be

appropriate to provide non-English language-based tests to students who receive their instruction in that language. For deaf and hard-of-hearing students who receive instruction in ASL, for example, presenting tests in ASL is a justifiable strategy for removing language barriers resulting from low English fluency. But we emphasize that such translation is only likely to benefit students who learned their subjects in ASL; having ASL as their first or primary language of discourse outside the classroom is not a sufficient condition to justify test presentation in ASL.

Adapting tests from one language to another is not a simple matter of word-by-word translation. Hambleton (2005) discusses many issues that need to be addressed in test adaptation, including culture and language differences, qualification of the translators, test administration variations, and similarity of curricula across classrooms for where the two languages are used. Studies on cross-language testing have been conducted to address various validity issues (e.g., Sireci, 1997; Hambleton, Merenda, & Spielberger, 2005). However, in the case of ASL, there is no published systematic research indicating how translation affects the structure and difficulty of test items. Officials in the Delaware Department of Education (Maihoff et al., 2000) conducted a small pilot study investigating how ASL translation influences test administration on mathematics tests presented using a television with DVD player format to deaf and hard-of-hearing students in two schools for the deaf in Delaware and Pennsylvania. The pilot study found that the students generally welcome the new testing format because they could better understand some mathematics concepts through DVD presentation. But due to limitations of testing instruments, this study was unable to examine the effect of DVD format on student performance. Johnson, Kimball, and Brown (2001) discussed challenges with the use of ASL as an accommodation for deaf and hard-of-hearing students in Washington State but did not answer the key question about how such accommodation impact the validity of assessment. Tindal (2006) has reported no discernible effect for tests of mathematics presented in ASL among students at schools for the deaf in the Pacific Northwest, where tests were presented using digital video displayed on a personal computer mon-

itor with point-and-click answer selection; however, no technical documentation is available to evaluate the quality of this investigation.

Given that there has been limited research on ASL adaptation of written English assessment instruments, we highlight a few of the concerns that must be addressed in future work. In ASL, the meaning of particular concepts can be presented by hand shapes or hand movements in ways that are not available in written English. For example, the position of the hand articulating a larger number may be higher than that for articulating a smaller number, especially when embedded in a comparative construct (e.g., comparing how many marbles John has to how many Sue has). Also, inherent to the grammar of ASL, repetition and use of space can depict a mathematical construct in a way that is awkward in English (e.g., the number sign of 4 may be repeated three times in slightly different locations in front of the body to indicate *four in each of the three groups*). Such “number cues” or “mapping cues” may potentially affect how problems are solved and, therefore, their level of difficulty (Ansell & Pagliaro, 2001, 2006; Kritzer, Pagliaro, & Ansell, 2004). Keep in mind, however, that nearly all the research to date has investigated elementary arithmetic story problems using relatively small integer quantities, and the language used at this elementary level does not generalize to arithmetic with larger numbers, fractions, decimals, etc. As a consequence, the availability of “cues” may not persist or, worse, become misleading at later stages of learning.

Another unique issue in ASL test adaptation is that it changes the nature of the examinees’ testing experience in that the items are presented in the everyday, face-to-face discourse mode of the classroom rather than in a denuded written form. This means that an ASL version may measure a construct that is different from what the original written test intended to measure, which creates potential threats to test comparability. Valid inferences may be drawn about what students know and can do, but comparisons with performances on written tests may not be valid. Validity studies are required to determine the extent to which any test performance difference is explained by the presentation change associated with ASL adaptation. That is, in addition to investigating whether the



manner in which the content of the test is communicated affects test performance (i.e., written text, which does not exist for signed languages, or face-to-face discourse), the test developers need to know whether language changes may affect the difficulty level of an item because, for example, an unusual, rare, or otherwise challenging word in English translates only into a common, everyday, and easily understood word in the target language.

Currently, researchers depend on the analogy with the “read aloud” test accommodation for hearing students to evaluate the prospects for ASL translation validity. Helwig and coworkers (Helwig, Rozek-Tedesco, & Tindal, 2002; Helwig & Tindal, 2003) have demonstrated that reading aloud test items accommodates the reading difficulties of hearing students who would otherwise have their performance artificially depressed relative to their true mathematical ability but does not artificially inflate the scores of poor readers, generally. Further, the read-aloud accommodation provides no benefit to good readers regardless of mathematics skill and ability. These results indicate that the face-to-face nature of the read-aloud accommodation does not significantly alter the test construct for hearing students, a conclusion which provides indirect support for using ASL test adaptation for deaf and hard-of-hearing students.

## Conclusion

In summary, deaf and hard-of-hearing student participation in large-scale assessment programs has undergone significant changes, though their test performance profile has changed little over the last three decades. The achievement gaps between deaf and hard-of-hearing students and their hearing peers remain large. It may well be that lack of opportunity to learn, which itself may have its roots in language acquisition delays (Sachs, Bard, & Johnson, 1981; Schick, De Villiers, P., De Villiers, J., and Hoffmeister, 2007) or the reported deficiency of curriculum and instruction in at least some programs serving deaf students (Kelly, Lang, & Pagliaro, 2003; Pagliaro & Ansell, 2002; Pagliaro & Kritzer, 2005), dominates any explanation of observed test performance. Nonetheless, causal attribution will remain elusive until un-

contaminated measurements are available. Valid and reliable instruments are necessary to disentangle the confounding factors that limit the ability of testing to reflect true academic achievement among deaf and hard-of-hearing students.

Lower costs and increased availability of computers and digital video equipment should facilitate research, development, and implementation of language-based accommodations. At the present time, beyond the early developments in test accommodation maintained by the Gallaudet Research Institute for the Stanford Achievement Test Series, only three states have made advances in standardized approaches to academic assessment accommodations for deaf and hard-of-hearing students: Virginia offers the same modified English test version to deaf and hard-of-hearing students as to ELLs; and Massachusetts and South Carolina offer video-based ASL presentations of their state tests. However, no reports on the psychometric properties of these statewide test accommodations are available for review.

The problem of standardized administration in Massachusetts, South Carolina, and Virginia has been addressed, but validity and reliability have not. When it comes to reliability, depending on the state population tested and the adaptation chosen, too few deaf and hard-of-hearing students participate in the state testing program in the same way at the same test level to obtain good estimates. And even if participation rates were high enough, no state has published reliability estimates for this subpopulation. When it comes to validity, certainly, these carefully adapted state tests have the same content validity as the written English versions, but construct validity does not have the same support. For example, in the absence of tests for ASL comprehension or ASL vocabulary, there is no way to develop the same kind of convergent and discriminant validity evidence that is obtained, for example, from observing that scores obtained from written tests of mathematics procedures and mathematics problem-solving correlate more strongly with each other than scores from either mathematics test with scores from tests of reading comprehension or reading vocabulary. Worse, both the written English and the ASL-adapted versions of these state tests lack studies providing evidence of their criterion-related validity (e.g.,



predicting course grades or knowledge and skills for successful employment, particularly in high-stakes regimes where diplomas are denied based on exit examination scores). Very little in the way of independent research activity has provided any evidence to fill this psychometric near vacuum for modified English or ASL-adapted state tests. Regardless of whether an “interaction effect” is observed for either the modified English text or video-based ASL presentation accommodations, extensive psychometric research employing experimental designs is urgently needed for future test development in this area. Otherwise, any large-scale testing program will continue to lack the scientifically based research support demanded by the same legislation that mandates test accommodations for students with disabilities or limited English proficiency. We urge practitioners and researchers to collaborate in the large-scale research efforts that are required to advance the development of valid, reliable, and effective assessments of academic achievement for deaf and hard-of-hearing students.

### Conflict of Interest

No conflicts of interest were reported.

### Acknowledgements

An earlier version of this paper was presented at the 2007 Annual Meeting of the American Educational Research Association, Chicago, Illinois.

### References

- Abedi, J. (2005). Issues and consequences for English language learners. In J. L. Herman & E. H. Haertel (Eds.), *104th Yearbook of the National Society for the Study of Education, Part II: Uses and misuses of data for educational accountability and improvement* (pp. 175–198). Malden, MA: Blackwell.
- Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74(1), 1–28. doi: 10.3102/00346543074001001.
- Abedi, J., Lord, C., & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance*. Los Angeles: UCLA Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learner's test performance. *Educational Measurement: Issues and Practice*, 19, 16–26. doi: 10.1111/j.1745-3992.2000.tb00034.x.
- Allen, T. E. (1984). *Out-of-level testing with the Stanford Achievement Test (7th ed.): A procedure for assigning students to the correct battery level*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Allen, T. E. (1986a). *Understanding the scores: Hearing-impaired students and the Stanford Achievement Test (7th Ed.)*. Washington, DC: Gallaudet Research Institute, Center for Assessment and Demographic Studies.
- Allen, T. E. (1986b). Patterns of academic achievement among hearing impaired students: 1974 and 1983. In A. N. Schildroth & M. A. Karchmer (Eds.), *Deaf children in America* (pp. 161–206). San Diego, CA: College-Hill Press.
- Allen, T. E., & Sligar, S. R. (1994). The assessment of deaf individuals in the context of rehabilitation. In R. C. Nowell & L. E. Marshak (Eds.), *Understanding deafness rehabilitation process* (pp. 113–142). Needham Heights, MA: Allyn and Bacon.
- Allen, T. E., White, C. S., & Karchmer, M. A. (1983). Issues in the development of special edition for hearing-impaired students of the seventh edition of the Stanford Achievement Test. *American Annals of the Deaf*, 128(1), 34–39.
- Ansell, E., & Pagliaro, C. (2001). Effects of a signed translation on the types and difficulties of arithmetic story problems. *Focus on Learning Problems in Mathematics*, 23(2, 3), 41–69.
- Ansell, E., & Pagliaro, C. M. (2006). The relative difficulty of signed arithmetic story problems for primary level deaf and hard-of-hearing students. *Journal of Deaf Studies and Deaf Education*, 11, 153–170. doi: 10.1093/deafed/enj030.
- Bello, M., Costello, P., & Recane, S. (2008). Participating in the Massachusetts Comprehensive Assessment System. In R. C. Johnson & R. E. Mitchell (Eds.), *Testing deaf students in an age of accountability* (pp. 181–193). Washington, DC: Gallaudet University Press.
- Blackorby, J., & Knokey, A.-M. (2006). *A national profile of students with hearing impairments in elementary and middle school: A special topic report from the special education elementary longitudinal study*. Menlo Park, CA: SRI International. Retrieved March 20, 2007, from the [http://www.seels.net/designdocs/HI\\_SPEC\\_TOPIC\\_112906\\_ww.pdf](http://www.seels.net/designdocs/HI_SPEC_TOPIC_112906_ww.pdf).
- Bloomquist, C. A. (1986). Minimum competency testing programs and hearing impaired students. In A. N. Schildroth & M. A. Karchmer (Eds.), *Deaf children in America* (pp. 207–229). San Diego, CA: College-Hill Press.
- Bosso, E. (2008). Testing, accountability, and equity for deaf students in Delaware. In R. C. Johnson & R. E. Mitchell (Eds.), *Testing deaf students in an age of accountability* (pp. 167–180). Washington, DC: Gallaudet University Press.
- California Department of Education. (2007). *Achievement gap for the deaf*. Supplement to State of Education Address delivered by Jack O'Connell, State Superintendent of Instruction, California Department of Education, Sacramento, CA,

- Retrieved October 3, 2010, from <http://www.cde.ca.gov/eo/in/se/agdeaf.asp>.
- Case, B. J. (2008). Accommodations to improve instruction and assessment of deaf students. In R. C. Johnson & R. E. Mitchell (Eds.), *Testing deaf students in an age of accountability* (pp. 51–62). Washington, DC: Gallaudet University Press.
- Cawthon, S. W. (2006). National survey of accommodations and alternate assessments for students who are deaf or hard of hearing in the United States. *Journal of Deaf Studies and Deaf Education*, 11, 337–359. doi: 10.1093/deafed/enj040.
- Cawthon, S. W. (2007). *Findings from the first two years of the National Survey of Assessment and Accommodations for Students who are Deaf or Hard of Hearing*. Paper presented in the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Cawthon, S. W. (2008). No Child Left Behind and schools for the deaf. In R. C. Johnson & R. E. Mitchell (Eds.), *Testing deaf students in an age of accountability* (pp. 92–114). Washington, DC: Gallaudet University Press.
- Cawthon, S. W., & Online Research Laboratory (2008). Accommodations use for statewide standardized assessments: Prevalence and recommendations for student who are deaf or heard of hearing. *Journal of Deaf Studies and Deaf Education*, 13, 55–96. doi: 10.1093/deafed/enm029.
- Cawthon, S. W., & Wurtz, K. A. (2009). Alternate assessment use with students who are deaf or hard of hearing: An exploratory mixed-methods analysis of portfolio, checklists, and out-of-level test formats. *Journal of Deaf Studies and Deaf Education*, 14, 155–177. doi: 10.1093/deafed/enn027.
- CCSSO SchoolMatters. (2010) Eighth grade English & math proficiency levels at Louisiana School for the Deaf grade 8. Retrieved October 3, 2010, from <http://www.schoolmatters.com/schools.aspx/q/page=sl/sid=27590/midx=GradeG8>.
- Clapper, A. T., Morse, A. B., Lazarus, S. S., Thompson, S. J., & Thurlow, M. L. (2005). *2003 state policies on assessment participation and accommodations for students with disabilities (Synthesis Report 56)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved August 31, 2007, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis56.html>.
- Emmorey, K., Bellugi, U., Frederici, A., & Horn, P. (1995). Effects of age of acquisition on grammatical sensitivity: Evidence from on-line and off-line tasks. *Applied Psycholinguistics*, 16(1), 1–23. doi:10.1017/S0142716400006391.
- Foster, C. (2008). One state's perspective on the appropriate inclusion of deaf students in large-scale assessment. In R. C. Johnson & R. E. Mitchell (Eds.), *Testing deaf students in an age of accountability* (pp. 115–135). Washington, DC: Gallaudet University Press.
- Gallaudet Research Institute. (1983). *Norm tables: 1982 Stanford Achievement Test, 7th Edition, Form E for use with hearing-impaired students*. Washington, DC: Gallaudet University.
- Gallaudet Research Institute. (1991). *Hearing impaired norm booklet: Stanford Achievement Test, 8th Edition, Form J*. Washington, DC: Gallaudet University.
- Gallaudet Research Institute. (1996). *Norms booklet for deaf and hard-of-hearing students: Stanford Achievement Test, 9th Edition, Form S*. Washington, DC: Gallaudet University.
- Gallaudet Research Institute. (2004). *Norms booklet for deaf and hard-of-hearing students: Stanford Achievement Test, 10th Edition, Form A*. Washington, DC: Gallaudet University.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum.
- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum.
- Helwig, R., Rozek-Tedesco, M. A., & Tindal, G. (2002). An oral versus a standard administration of a large-scale mathematics test. *Journal of Special Education*, 36, 39–47. doi: 10.1177/00224669020360010401.
- Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children*, 69, 211–225.
- Holt, J. A., & Allen, T. E. (1989). The effects of schools and their curricula on the reading and mathematics achievement of hearing impaired students. *International Journal of Educational Research*, 13, 547–562. doi:10.1016/0883-0355(89)90061-X.
- Holt, J. A., Traxler, C. B., & Allen, T. E. (1992). *Interpreting the scores: A user's guide to the 8th Edition Stanford Achievement Test for educators of deaf and hard of hearing students (Gallaudet Research Institute Technical Report 92-1)*. Washington, DC: Gallaudet University.
- Holt, J. A., Traxler, C. B., & Allen, T. E. (1997). *Interpreting the scores: A user's guide to the 9th Edition Stanford Achievement Test for educators of deaf and hard of hearing students (Gallaudet Research Institute Technical Report 97-1)*. Washington, DC: Gallaudet University.
- Individuals with Disabilities Education Act Amendments of 1997 (IDEA '97), Pub. L. No., 105-17 111 Stat. 37 *et seq.* (1997).
- Individuals with Disabilities Education Improvement Act of 2004 (IDEA '04), Pub. L. No., 108-446 118 Stat. 2647 *et seq.* (2004).
- Jensema, C. J., Schildroth, A. N., & O'Rourke, S. (1974). Score conversion tables and age-based percentile norms. Washington, DC: Gallaudet College, Office of Demographic Studies.
- Johnson, R. E., Liddell, S. K., & Erting, C. J. (1989). Unlocking the curriculum: Principles for achieving access in deaf education (Working Paper No. 89-3). Washington, DC: Gallaudet Research Institute, Gallaudet University.
- Johnson, E., Kimball, K., & Brown, S. O. (2001). American sign language as an accommodation during standard-based assessments. *Assessment for Effective Intervention*, 26, 39–47. doi: 10.1177/073724770102600207.
- Jones, M. (2008). Implications and complications of including deaf students in statewide assessments in Illinois. In R. C. Johnson & R. E. Mitchell (Eds.), *Testing deaf students in an*

- age of accountability (pp. 149–166). Washington, DC: Gallaudet University Press.
- Karchmer, M. A., & Mitchell, R. E. (2003). Demographic and achievement characteristics of deaf and hard of hearing students. In M. Marschark & P. E. Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (pp. 21–37). New York, NY: Oxford University Press.
- Kelly, L. P., & Barac-Cikoja, D. (2007). The comprehension of skilled deaf readers: The roles of word recognitions and other potentially critical aspects of competence. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 244–280). New York, NY: The Guilford Press.
- Kelly, R. R., Lang, H. G., & Pagliaro, C. M. (2003). Mathematics word problem solving for deaf students: A survey of practices in grades 6–12. *Journal Deaf Studies & Deaf Education*, 8, 104–119. doi: 10.1093/deafed/eng007.
- Kritzer, K. L., Pagliaro, C. M., & Ansell, E. (2004). Deaf/hard-of-hearing students' use of language cues to solve signed story problems. *EdPitt*, 1(1). Retrieved March 20, 2007, from <http://www.pitt.edu/~edpitt/currentissueK-RITZER.html>.
- Lazarus, S. S., Thurlow, M. L., Lail, K. E., Eisenbraun, K. D., & Kato, K. (2006). 2005 state policies on assessment participation and accommodations for students with disabilities (*Synthesis Report 64*). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved March 20, 2007, from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis64/>.
- Maihoff, N. A., Bosso, E., Zhang, L., Fischgrund, J., Schulz, J., Carlson, J., & Carlson, J. E. (2000). *The effects of administering an ASL signed standardized test via DVD player/television and by paper-and-pencil: A pilot study*. Dover, DE: Delaware Department of Education. Retrieved May 10, 2011, from <http://www.doe.k12.de.us/aab/files/Atch9.pdf>.
- Marschark, M., Lang, H. G., & Albertini, J. A. (2002). *Educating deaf students: From research to practice* (pp. 73–80). New York, NY: Oxford University Press.
- Mayberry, R. I., & Lock, E. (1998). *Critical period effects on grammatical processing: Privileged status of the first language. Paper presented at the American Psychological Society Convention, Washington, DC*.
- Mazzeo, C. (2001). Frameworks of state: Assessment policy in historical perspective. *Teachers College Record*, 103, 367–397. doi: 10.1111/0161-4681.00119.
- Meadow-Orlans, K. P., Mertens, D. M., & Sass-Lehrer, M. A. (2003). *Parents and their deaf children: The early years*. Washington, DC: Gallaudet University Press.
- Mitchell, R. E. (2004). National profile of deaf and hard of hearing students in special education from weighted survey results. *American Annals of the Deaf*, 149, 336–349. doi: 10.1353/aad.2005.0004.
- Mitchell, R. E. (2008). Academic achievement of deaf students. In R. C. Johnson, R. E. Mitchell (Eds.), *Testing deaf students in an age of accountability* (pp. 38–50). Washington, DC: Gallaudet University Press.
- Mitchell, R. E., & Karchmer, M. A. (2004a). When parents are deaf vs. hard of hearing: Patterns of sign use and school placement of deaf and hard-of-hearing children. *Journal of Deaf Studies and Deaf Education*, 9, 133–152. doi: 10.1093/deafed/enh017.
- Mitchell, R. E., & Karchmer, M. A. (2004b). Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies*, 4, 138–163. doi: 10.1353/sls.2004.0005.
- Mitchell, R. E., & Karchmer, M. A. (2005). Parental hearing status and signing among deaf and hard of hearing students. *Sign Language Studies*, 5, 83–96. doi: 10.1353/sls.2005.0004.
- Mitchell, R. E., Qi, S., & Traxler, C. B. (2007). *Stanford Achievement Test, 10th Edition. National performance norms for deaf and hard of hearing students: A technical report*. Unpublished manuscript. Washington, DC: Gallaudet Research Institute, Gallaudet University.
- Moore, P. (2008). The potential harm to deaf students of high-stakes testing in California. In R. C. Johnson & R. E. Mitchell (Eds.), *Testing deaf students in an age of accountability* (pp. 217–238). Washington, DC: Gallaudet University Press.
- Morford, J. P., & Mayberry, R. I. (2000). A reexamination of “early exposure” and its implications for language acquisition by eye. In C. Chamberlain, J. P. Morford, & R. I. Mayberry (Eds.), *Language acquisition by eye* (pp. 111–127). Mahwah, NJ: Lawrence Erlbaum.
- Mowl, H. M. (1985). *Performance of deaf students on a standard and modified minimum competency test*. (Unpublished doctoral dissertation), University of Pittsburgh, Pittsburgh, PA; *Dissertation Abstracts International*, 46, 3685.
- National Research Council. (2005). Hearing loss: Determining eligibility for social security benefits. Committee on Disability Determination for Individuals with Hearing Impairments. In Robert A. Dobie & Susan B. Van Hemel (Eds.), Board on Behavioral, Cognitive and Sensory Sciences, Division of Behavioral and Social Sciences and Education (pp. 180–189). Washington, DC: The National Academies Press.
- No Child Left Behind Act of 2001 (NCLB), Pub. L. No., 107-110 115 Stat.1425 et seq. (2002).
- Office of Demographic Studies. (1969). *Academic achievement test performance of hearing impaired students*. Washington, DC: Gallaudet College.
- Pagliaro, C. M., & Ansell, E. (2002). Story problems in the deaf education classroom: Frequency and mode of presentation. *Journal of Deaf Studies and Deaf Education*, 7, 107–119. doi: 10.1093/deafed/7.2.107.
- Pagliaro, C. M., & Kritzer, K. L. (2005). Discrete mathematics in deaf education: A survey of teachers' knowledge and use. *American Annals of the Deaf*, 150, 251–259. doi: 10.1353/aad.2005.0033.
- Parent Resource Center Fairfax County Public Schools. (2004). New information about standard assessments and students with disabilities. *Special Edition: A Newsletter for Parents of Students Who Receive Special Education Services*, Winter(20). Retrieved November 1, 2005, from [http://www.fcps.edu/ss/prc/newsletters/SpringNewsletter\\_2004.pdf](http://www.fcps.edu/ss/prc/newsletters/SpringNewsletter_2004.pdf)
- Sachs, J., Bard, B., & Johnson, M. L. (1981). Language learning with restricted input: Case studies of two hearing children

- of deaf parents. *Applied Psycholinguistics*, 2, 33–54. doi:10.1017/S0142716400000643.
- Schick, B., De Villiers, P., De Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *Child Development*, 78, 376–396. doi: 10.1111/j.1467-8624.2007.01004.x.
- SchoolDigger.com. (2010). *Louisiana School for the Deaf*. Retrieved October 3, 2010, from <http://www.schooldigger.com/go/LA/schools/0001801789/school.aspx?entity=60&grade=10>.
- Sireci, S. G. (1997). Problems and issues in linking tests across languages. *Educational Measurement: Issues and Practice*, 16(1), 12–19. doi: 10.1111/j.1745-3992.1997.tb00581.x.
- Sireci, S. G., Scarpato, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75, 457–490. doi: 10.3102/00346543075004457.
- South Carolina Department of Education. (2005). *High School Assessment Program Test Administration Manual—Spring 2005*. Retrieved August, 2 2006, from <http://www.myschools.com/offices/assessment/Programs/SWD/HSAPcS05.pdf>.
- Steffan, R. C., Jr. (2008). The effect of No Child Left Behind at the Maryland School for the Deaf and nationwide. In R. C. Johnson, R. E. Mitchell (Eds.), *Testing deaf students in an age of accountability* (pp. 204–216). Washington, DC: Gallaudet University Press.
- Thurlow, M. L., Johnstone, C., Thompson, S. J., & Case, B. J. (2008). Using universal design research and perspectives to increase the validity of scores on large-scale assessments. In R. C. Johnson, R. E. Mitchell (Eds.), *Testing deaf students in an age of accountability* (pp. 63–75). Washington, DC: Gallaudet University Press.
- Tindal, G. (2006). *Accommodations research: Reconsidering the test accommodation validation process: A paradigm for research design with initial outcomes*. Presentation at the ETS conference “Accommodating Students With Disabilities on State Assessments: What Works? Savannah, GA.
- Trybus, R. J., & Karchmer, M. A. (1977). School achievement scores of hearing impaired children: National data on achievement status and growth patterns. *American Annals of the Deaf Directory of Programs and Services*, 122, 62–69.
- U.S. Department of Education. (2005) Proposed rules: 34 CFR Parts 200 and 300 Title I—Improving the academic achievement of the disadvantaged; Individuals With Disabilities Education Act (IDEA)—Assistance to states for the education of children with disabilities. *Federal Register*, 70, 74624–74638.