Afterword

Ross E. Mitchell
and Robert C. Johnson

It is our sincere hope that this volume will serve to inform and enlighten readers about the many challenges that must be overcome to make a system of test-based accountability workable and beneficial for deaf students and their schools. We appreciate the fact that the No Child Left Behind Act is especially concerned with ensuring that groups of students historically educated outside the mainstream, as nearly all deaf students were until 1975, be given the best possible chance to attain proficiency in a general curriculum. At the same time, the expectation that deaf students (with the exception of those who are also severely cognitively impaired) must demonstrate their proficiency on English-based tests strikes us as problematic at best, considering the difficulties these students usually experience learning the printed code of a language they cannot hear.

One major reason politicians value testing is because externally imposed tests are supposed to eliminate an information asymmetry between the public (including policymakers) and schools (McDonnell, 2005). In other words, in the absence of testing, schools have monopoly control over such indicators of student learning as grades and, therefore, cannot be held accountable for how well or poorly students *really* are learning. But as Michael Jones (this volume) documents at length, politicians unfortunately rarely understand the complex on-the-ground reality of test-based accountability for deaf students. And with the exception of states like South Carolina (Foster, this volume), state departments of education have not fully explored and developed approaches to assessment that are standards-based, technically strong, and reflective of educational practices appropriate for this and other special populations.

There is no doubt that state and federal test-based accountability has transformed public schooling, including deaf education, over the last two decades. As several state, district, and school officials have noted, the general curriculum is no longer only for hearing students; deaf students are to receive instruction in the general curriculum and must be prepared for standards-based tests as well. Moreover, teachers are to identify "entry points" into the general curriculum for students with additional disabilities (Bello et al., this volume). Though not necessarily with the same quality or consistency from state to state, a tremendous effort has been and continues to be made to fit deaf education into state and federal accountability systems.

In other words, there has been tremendous progress within schools in creating the *opportunity to learn* the material that is supposed to be learned in a system of test-based accountability. However, as we address later, some significant opportu-

nity to learn issues still loom large. Also, true accountability assumes that the *opportunity to demonstrate learning* is valid and reliable, but there are legitimate reasons to doubt that such inferences from test performance necessarily hold true in the case of deaf students. Also, for accountability to serve the development of improved curriculum and instruction, assessment needs to provide an *opportunity to diagnose and treat learning difficulties.* We shall close with a brief comment on how all systems of test-based accountability currently fail to provide this final opportunity and how this problem might be rectified.

The fit between deaf education and test-based accountability systems is still imperfect. The first major source of disjuncture we address is the testing itself, which, at its worst, fails to provide the opportunity to demonstrate learning, and, at its best, leaves considerable room for improved opportunities to demonstrate learning. The second disjuncture arises in the consideration of when and how to hold schools accountable for deaf students' learning; there are still important opportunity-to-learn issues in the delivery of special education to deaf students. The final disjuncture, which afflicts the entire system of accountability, regardless of deafness among students, is the information quality derived from accountability assessments. The information density provided by a student's responses to test items is too low to provide an adequate diagnosis of the student's learning difficulties, let alone to help determine how instruction should be modified to remedy the problem manifested by erroneous answers.

In the case of the testing itself, the assumptions underlying most standards-based and standardized tests simply do not hold true for deaf students. The most important failure is the assumption of comprehensibility. Tests are developed in English on the presumption that, even by the third grade, most students are reasonably fluent in the language and are able to associate written words with spoken words. In no case is there a state, district, or school that provides evidence for public accountability based on administering a direct assessment of deaf students in American Sign Language (ASL), the language in which the students who would take such a test ostensibly receive their instruction and through which they communicate their understanding. Instead, the best options offered to date (and not offered in most states), are when English language tests are carefully and uniformly translated into ASL and, when cognitively inaccessible or patently inappropriate items are present (e.g., asking deaf students to identify sounds typically heard at a beach), substitute items measuring the same construct are found or developed to overcome unmistakable biases.

Of course, English reading and writing fluency is a target outcome for deaf students as well as hearing students, but it is the fundamental lack of ready access to English, an inaccessibility more profound than that experienced by non-English-speaking immigrant children, that gives rise to this special education category. We wish to emphasize that in the absence of cognitive disabilities, it is not the capacity for abstract or sophisticated thought that is disabled by deafness but the opportunity to acquire English and all that is communicated through English with the same access and ease as those who can hear.

The absence of direct assessment in ASL is what gives rise to several of the test

administration burdens and debates that surround deaf students' participation in accountability testing. Translation of an assessment instrument from its original source language to a target language (in this case, English to ASL) is guaranteed to introduce difficulties that may be avoided by developing both language versions simultaneously (see Tanzer, 2005). A parallel ASL assessment presented in a visual medium (e.g., videotape, DVD, or streaming video) would impose no additional preparation time for staff. Also, there would be no added threats to test security, no fear of incompetent or idiosyncratic signed interpretations, and no need for interpreter guidelines. Skilled personnel would not have to be reassigned to present the test. There would be no need for a bilingual dictionary or thesaurus. Extended time and other accommodations related to processing an English-based test would no longer be required, though an assessment in ASL might still require more time to administer. Finally, by developing ASL and English tests in parallel, ridiculous expectations about item equivalence between languages could be overcome. In particular, visual concepts like parallel lines and geometric shapes become recognizable as vocabulary rather than mathematics (e.g., "identify which shape is a triangle" is visual vocabulary; "identify which shape is a plane figure for which its interior angles sum to 180 degrees" is mathematics).

At the same time, a direct assessment in ASL does not eliminate all identified concerns. When it comes to tests of English language proficiency itself (e.g., reading comprehension), students who cannot respond in written English or identify the correct English word(s) or sentence(s) among a set of response choices find that they may require the assistance of a dictionary, thesaurus, interpreter, or scribe. Clearly, in cases such as these, getting at what these students know must happen in a manner that may differ substantially from any standardized assessment protocol. Often, the need for response supports at this level is viewed as modification that—even if the logistics of cost, personnel, equipment, space, and time were manageable—would invalidate test scores. This creates a peculiar and corrosive condition wherein getting the best access to what students know is not legitimate in the eyes of the state, while participating in an unmodified and state-approved testing condition can be a worthless assessment exercise (i.e., one that provides essentially no information about student learning) for students, teachers, and schools alike, not to mention parents and the greater public.

When and how to hold schools accountable for deaf students' learning remains a problem regardless of how test development and implementation issues are resolved. The emphasis on the unique needs and circumstances of the child receiving special education, formalized in the Individualized Education Program (IEP), creates an added layer of complexity for schools and programs serving deaf students. In particular, an agreed-upon plan that is purported to meet the needs of the deaf child may, in fact, be more strongly aligned with the interests of the child's parents, teachers, and school administrators than with an unbiased assessment of the child's needs.

Even allowing for substantial uncertainty in the best initial course of action, once it is recognized that the IEP has failed, the deaf child is as likely to be transferred to a new school capable of delivering an alternative program of instruc-

tion—often a state school for the deaf—as to have a substantial IEP revision enacted at the child's current school. When this move occurs, years of capitalized failure are transferred to the receiving school. The probability that the receiving school can overcome the early failure in a year is quite low. The likelihood that the receiving school will effectively bring the child up to a proficient level before reaching the maximum age allowable—assuming the student will persist rather than drop out or take a certificate of completion in lieu of a high school diploma—is unknown. In other words, school accountability for students in special education, especially if these students are mobile, should be considered very carefully. Attributing failure to a school or program can be a rather murky problem that is not justly handled by setting arbitrary timelines for transfer of responsibility for student performance. Just as it is not sensible to absolve the sending school of all responsibility and leave the receiving school alone to shoulder the blame for a student's poor performance, it is similarly not reasonable to perpetually hold the sending school responsible and leave the receiving school free of any formal accountability (see Bosso; Cawthon; Fischgrund; Jones; Steffan, this volume).

Finally, even if states were to create ways to get the results from summative assessments back to parents and teachers in time to inform instruction plans for the following year, the information density would be too low to facilitate effective planning. As M. David Miller (2007) has demonstrated, even the very best achievement tests ever developed would have to be doubled in length relative to current testing protocols to provide instructionally useful feedback. And given the fact that many accountability assessments are not of this premium quality (i.e., their reliability is still good, but not the very highest), these standards-based tests might have to be as much as ten times longer to provide such information.

In order to address the inadequacy of information in current accountability assessment systems for instructional purposes, we suggest that the secure test approach be abandoned in exchange for overwhelmingly large and public test item pools. That is, since it is very unlikely that any school or program will cover every possible spelling word, mathematics problem, novel, short story, science problem, or history topic in its state-adopted or state-approved curriculum materials, and since these words, problems, stories, topics, etc., should serve as legitimate items on standards-based tests, it should be similarly possible to create an overwhelmingly large pool from which to randomly select items for each year's assessment. Of course, there are technical challenges related to establishing reliability and comparable difficulty, especially in states that wish to have their own uniquely designed assessment systems (except for the largest states, it would be difficult to establish item parameters for such a large item pool), but there would be no doubt about what would possibly appear on the state test. Representative items could be used for assessment throughout the year, and the feedback from these intermediate results could be trusted to fairly correspond with the learning and performance requirements of the year-end summative accountability test. Though the idea of national tests does not appeal to most states, when it comes to special populations, especially low-incidence populations like deaf students, it is only at

this scale that item characteristics can be examined to ensure that test validity and reliability hold true for *all* students who participate in statewide testing programs.

Other solutions to assessment dilemmas posed by special student populations will surely be proposed and experimented with as long as statewide or national assessments continue to be used to measure achievement levels in this age of accountability. Many chapters in this book contain hints of possible solutions to which we have added some in this afterword. But attention should also be given here to the importance of nurturing reciprocal rather than adversarial relationships between administrators of schools serving deaf students and staff in state departments of education.

In some states, individuals charged with implementing assessment programs appear to be driven by preconceptions that do not always square with the experience of educators familiar with special populations. A striking example of this was presented by Jon Levy at the 2002 Gallaudet University conference concerning high stakes testing and deaf students. Levy, principal of a regional program for deaf students at University High School in Orange County, California, said that during a workshop with representatives of California's State Department of Education he had informed the officials that students in his program were reading at fifth- to seventh-grade levels. This was significantly higher than national norms for deaf students, and he had reason to be proud of this statistic. He had then expressed concern that a new mandate requiring that all students in California pass tests in English and math at a tenth-grade level might prevent 80 to 85% of California's deaf students from receiving high school diplomas, greatly reducing their options for future employment or postsecondary education. Levy described the officials' reaction this way: "The response was clear, stark, and very upsetting. They said to me that it is you administrators and your teachers of the deaf who are at fault. If you simply raised your standards and had higher expectations for these deaf children, they would be reading at the twelfth-grade level and passing the High School Exit Exam" (Johnson, 2003, p. 7).

We concur with Pat Moore (this volume) and testing scholar Stuart Yeh (see, for example, Yeh, 2008) that some states, like California, need to adopt a tiered diploma system to provide proper recognition of and incentives for the highest academic achievements attainable by *all* students, including deaf students. As reviewed by Yeh, many standards-based high-stakes exit exams have little or no external validity; that is, much of their test content is irrelevant to the knowledge and training needs of employers who require a high school diploma for employment eligibility. More importantly, such exit exams do not play the incentive role policymakers intended. The quantum leap from minimum competency testing to world-class standards appears to have taken away the incentive for low achievers to meet the performance mark consistent with the social and economic meaning of the high school diploma in this country. All of the growth and leading success in employment and higher education among deaf students reported from the National Longitudinal Transition Study 2 (Wagner, Newman, Cameto, & Levine, 2005) may stall or reverse if current policies are left unchanged in several states.

Although some of our chapter authors, professionally involved in deaf education for decades, express deep frustrations with a lack of understanding among government officials in their state, this book contains signs of hope that assertive, persistent, well-meaning efforts by school personnel can lead to constructive working relationships with state departments of education, fostering collegial exchanges and joint efforts to make assessments of deaf students fairer and more valid (see especially Jones; Bosso; Bello, Costello, & Recane; Fischgrund; and Foster, this volume). We sincerely hope that state and federal officials will read this book and do their part to open a dialogue with experienced teachers and administrators in schools or programs for deaf students to learn what those who understand this student population's difficulties have to say about policies, procedures, and test forms that are not working well. We hope that astute readers can pick up many ideas here concerning work that is being carried out in some states that is helping make the curriculum and tests more accessible to deaf students—hence, fairer and more valid for measuring what deaf students do or do not know.

REFERENCES

Johnson, R. C. (2003). High stakes testing conference held at Gallaudet University. *Research at Gallaudet, 2003* (Fall), 1–11.

McDonnell, L. (2005). Assessment and accountability from the policymaker's perspective. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement: The 104th yearbook of the National Society for the Study of Education,* Part 2 (pp. 35–54). Chicago: National Society for the Study of Education.

Miller, M. D. (2007, January). *School improvement and educational accountability.* Paper presented at the 2007 CRESST Conference: The Future of Test-based Educational Accountability, University of California, Los Angeles.

Tanzer, N. K. (2005). Developing tests for use in multiple languages and cultures: A plea for simultaneous development. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 235–264). Mahwah, NJ: Lawrence Erlbaum.

Wagner, M., Newman, L., Cameto, R., & Levine, P. (2005). *Changes over time in the early postschool outcomes of youth with disabilities.* A report of findings from the National Longitudinal Transition Study (NLTS) and the National Longitudinal Transition Study-2 (NLTS2). Menlo Park, CA: SRI International.

Yeh, S. S. (2008, January 28). Raising the bar and reducing failure on state-mandated exit exams. (ID Number: 14939). *Teachers College Record.* Retrieved April 23, 2008, from http://www.tcrecord.org/PrintContent.asp?ContentID=14939.