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

Stephanie W. Cawthon
The University of Texas at Austin
Keith A. Wurtz
Walden University

The purpose of this paper is to present findings on alternate assessments for students who are deaf or hard of hearing (SDHH). Drawn from the results of the “Second National Survey of Assessments and Accommodations for Students Who Are Deaf or Hard of Hearing,” this study investigated three alternate assessment formats: portfolio, checklists, and out-of-level testing. Analysis includes descriptive data of alternate assessment use across all three formats, qualitative analyses of teacher perspectives, and an exploratory logistic regression analysis on predictors of alternate assessment use. This exploratory analysis looks at predictors such as state policy, educational setting, grades served, language of instruction, and participant perspectives. Results indicate that predictors at the student, teacher, and system level may influence alternate assessment use for SDHH.

In the current high-stakes testing environment, large-scale, standardized assessments are the primary way that states measure student achievement (e.g., No Child Left Behind Act [2002]). When including students with disabilities in large-scale assessments, additional considerations need to be made as to how students will access the assessment content (Individuals with Disabilities Education Act, IDEA, 1997). Students with disabilities may face barriers to the content of the assessment for many reasons, including the paper-and-pencil format used by many test vendors. In some cases, testing accommodations can remove the barriers to accessing the target skill of the assessment and allow students to meaningfully participate in the test (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Bolt & Thurlow, 2004). In cases where effective accommodations are not available, or changes made with the accommodation put the validity of the test at risk, “alternate assessment” formats can be used to effectively measure student knowledge and skill.

The purpose of this paper is to discuss alternate assessment use for students who are deaf or hard of hearing (SDHH). This manuscript will focus on three common alternate assessment formats selected because of their use with this student population (Cawthon & the Online Research Lab, 2006) or their inclusion in state alternate assessment policy (Thompson, Johnstone, Thurlow, & Altman, 2005): portfolios of student work, checklists of student knowledge and skill, and out-of-level testing. This literature review will first discuss the purpose of alternate assessment use, followed by a discussion of the format and technical issues for each of the three targeted alternate assessment formats. The paper will then turn to what is known about alternate assessment use for SDHH and the guiding research questions for this study. The paper then provides alternate assessment results from the “Second Annual National Survey of Assessment and Accommodations for Students Who Are Deaf or Hard of Hearing.” The discussion section considers the implications of these results for SDHH and their inclusion in the No Child Left Behind (NCLB) accountability framework.
Alternate Assessment

Alternate assessment formats. Alternate assessments allow students who cannot participate in standardized assessments (even with accommodations) to be included in the large-scale evaluation process. These students are usually identified as those who have significant cognitive disabilities (CDs), though that definition and its application varies from state to state (Thompson et al., 2005). It is important to note that many students whose primary disability is deaf or hard of hearing also have other disabilities. Approximately 40% of students counted in the 2003–2004 Gallaudet Annual Survey were listed as having an additional disability (Gallaudet Research Institute [GRI], 2005), including learning disabilities (LDs), cerebral palsy (CP), mental retardation, emotional disturbance, and attention deficit disorder. Although each individual disability may not be considered severe, a student with multiple disabilities may face challenges in participating in assessments that one with only hearing loss may not.

State guidelines for developing alternate assessments have occurred later than standardized tests and accommodations (Koretz & Barton, 2003; Quenemoen, Rigney, & Thurlow, 2002). Development of assessments for students in regular education has had at least 50 years to develop into a robust field of study; this timeline has been compressed and accelerated for developing assessments for students with complex disabilities, due in part to the IDEA and NCLB legislation. Many states are still revising their alternate assessment policies in response to changes in federal laws, transitioning from a focus on functional skills to those that reflect curriculum standards for all students (Thompson et al., 2005). In the development of their assessment systems, states may choose from a range of alternate assessment strategies for students with the most significant disabilities. A brief review of the three formats of alternate assessments explored in this study, followed by a discussion of reliability and validity issues, is given below.

Portfolio (body of evidence). A portfolio approach to alternate assessment typically involves gathering together artifacts and documents from the student’s work in the classroom to be used as a collection of evidence of student’s content knowledge and skill (Turner, Baldwin, Kleinert, & Kearns, 2000; Wiener, 2006). By 2005, approximately half of all states (including the District of Columbia) (n = 25) used this approach in their alternate assessments (Thompson et al., 2005). Of these states, 13 evaluated students based on a standardized set of items. For example, in Kentucky, a review board of three teachers scored a collection of work from multiple subject areas that span the work done in the previous 3 years. The teachers used a rubric including how the student met state academic standards, made progress toward Individualized Education Program (IEP) objectives, can apply skills to multiple settings, and whether the student had opportunities to develop relationships with peers (Kentucky Department of Education, 2007).

Checklist of knowledge and skill. The second format of alternate assessments explored in this study is a checklist of student knowledge and skills. Approximately a third of all states (30%) used a checklist in 2004 assessments, but this number decreased to only 14% (n = 7) in 2005, the year of the data collected in the current study (Thompson et al., 2005). The checklist consists of a summary of a student’s competencies that align to state standards and expectations for student achievement. The checklist can be a simple “yes” or “no” of targeted skills or a rating scale of relative progress that a student has made toward skill mastery. Compared with the portfolio, a checklist is not scored by an outside party and can, in some cases, be submitted without student work.

Out-of-level testing. Out-of-level testing is different from the other formats of alternate assessments. By definition, it measures alternate standards rather than grade-level standards. Although the test is in a standardized format, out-of-level testing is considered to be an alternate assessment because scores cannot be compared with those of their grade level, typically developing peers. It is also one of the least commonly allowed formats in state assessment policies, with fewer states using out-of-level testing than even just 4 years prior (VanGetson, Minnema, & Thurlow, 2004). At the time of the study (2003–2004), 13 states had some form of out-of-level testing, but six of these
states had pending changes reducing its use in state-wide assessment. In their analysis, VanGetson et al. (2004) examined published testing policies for each state for statements regarding out-of-level testing options (either allowed or prohibited). Authors then provide further descriptions of the context for allowed out-of-level testing, including test subject, test purpose, and student selection criteria. These findings were corroborated with discussion with state administrators in the annual Survey of State Directors of Special Education conducted by the National Center for Education Outcomes (e.g., Thompson & Thurlow, 2003). One rationale behind an out-of-level test is that it can help prevent the significant frustration experienced by a student when he or she takes a test that is far above their academic proficiency. It may also be more meaningful to teachers and parents to test students on the content of instruction instead of material that they have not had access to in the classroom.

From a technical standpoint, the out-of-level tests are easier to score than portfolios and do not have some of the reliability issues that come with subjective evaluations of student work.

The use of subgroup norms for a specific population, in this case, for SDHH, is useful in understanding how a student is achieving relative to their peers. Researchers at the Gallaudet Research Institute (GRI) have compiled an extensive database on student achievement scores from the Stanford Achievement Tests (SAT), from the 7th edition through the present 10th edition (GRI, 1983, 1991, 1996, 2004). As part of the norming process, students take an out-of-level “screening test” to determine at what level they should participate in the standardized assessment. In other words, the norming process involves participation in an out-of-level test (Qi & Mitchell, 2007). This history may have an impact on state standardized assessment strategies used with SDHH.

There are significant limitations, however, in use of out-of-level testing in accountability reforms such as NCLB. First, out-of-level testing used for state standardized assessments is different than the process for developing deaf or hard of hearing norms on the SAT. Student scores are not compared against a different norm, or “average” level for a subgroup, but against those for the student group as a whole. By extension, out-of-level testing means that the score can no longer be treated as measuring the same skills as standardized test scores for the student’s grade-level peers. In other words, the score from a student taking a test below grade level cannot be compared against those from students taking the on-grade level exam. For states that do allow out-of-level testing, the default is to assign a failing grade to the score regardless of how well a student performs on the test. Depending on the state policy and data reporting practice, passing scores on an out-of-level test may be reported as “not proficient” or “below proficient” (Minnema, Thurlow, Bielinski, & Scott, 2000). This practice leads to a reduction in information about knowledge and skills the student may have, even if it is below grade level. Out-of-level testing thus has a limited role in assessment used to evaluate student progress toward grade-level standards.

Reliability and validity. Issues of reliability and validity are especially challenging for alternate assessments. One reason for this is the central role of the person who compiles or assembles the evidence for student proficiency. “Reliability” refers to the consistency of assessment ratings, both across participants and across evaluators. For example, in an effort to maintain high-quality standards and reliable results, all teachers who submit portfolios in Kentucky must also attend workshops that explain how alternate assessments are scored. This training helps ensure that a student’s portfolio is correctly developed and compiled, thereby avoiding penalization due to missing or under developed items. This training also encourages teachers to provide students with opportunities to learn the material that is aligned to state standards and focused on their IEP goals. There is little information available as to the extent of this process across states.

Work on the “ validity” of alternate assessments focuses mainly on the portfolio approach (Johnson & Arnold, 2004). The validity of a student’s score can be conceptualized in a number of ways. One question is whether the alternate assessment is aligned to state academic content standards, as required by NCLB (Roach, Elliott, & Webb, 2005). Sufficient alignment, or content validity, is important because the purpose
of both standardized and alternate assessments is to measure student achievement on grade-level standards. Many states have recently revised their alternate assessment systems to meet these new guidelines, subject to peer review by the Department of Education (Thompson & Thurlow, 2003). States are responsible for developing their own evaluations of test validity as part of this process. Public reports of alternate assessment validity studies are thus only available for a handful of states (e.g., Johnson & Arnold, 2004; Roach et al. 2005; Turner, Baldwin, Kleinert, & Kearns, 2000; Weiner, 2006; Yovanoff & Tindal, 2007). Issues raised in these studies include (a) number of skills used to substantiate a student’s performance on standards; (b) range of standards covered by the assessment; (c) depth of knowledge measured; (d) use of single skill to “count” for multiple criteria; (e) equating proficiency ratings with scores from standardized assessments; and (f) variance in the above issues across subject areas (e.g., reading, math, science). State assessments vary in the extent to which they meet these technical quality criteria. Even with extensive teacher training, alternate assessment implementation requires significant investment in rigorous test design and assessment validation (Johnson & Arnold, 2004).

Alternate Assessment and NCLB

NCLB places significant restrictions on the proportion of student scores from alternate assessments that can be used toward a district’s calculation of proficiency (Yell, Katsiyannas, & Shiner, 2006). Only a very small proportion (initially 1%) of students can be included in an alternate assessment, typically those with the most severe CDs. Scores of alternate assessments above these cutoffs for alternate assessment use are counted for participation, but not toward school, district, and state proficiency rates. Recent changes to the legislation allow up to an additional 2% of a school’s eligible test takers to participate in alternate assessments that measure performance on “modified academic achievement standards” (Federal Register, December 15, 2005; Federal Register, April 9, 2007). The purpose of this new category is to recognize that some students with mild to moderate disabilities will not meet grade-level IEP proficiency goals and may not be able to meaningfully participate in a standardized assessment. However, these regulations only come into effect when it can be verified that students (a) received appropriate accommodations and (b) had the opportunity to learn grade-level content. In developing modified standards, states may change the level of mastery required to meet grade-level standards, but the content areas must remain the same. These new regulations are meant to give states further flexibility in what students are expected to learn and how their performance is measured. Furthermore, states can report student proficiency on modified standards in their overall evaluation of school effectiveness and progress toward NCLB Adequate Yearly Progress measures.

Alternate Assessments and SDHH

The first study by the authors, the “First Annual Survey of Assessments and Accommodations for Students Who Are Deaf or Hard of Hearing” provided a preliminary look at nationwide trends in alternate assessment use with SDHH (Cawthon, 2006). The unit of analysis for this initial study was the “school or program” that served SDHH. A total of 71 schools or programs had students who participated in alternate assessments. A total of 45% of schools or programs reported using an out-of-level test for at least one student, 42% a work sample, 37% used a curriculum-based assessment, 24% used a checklist or structured observations, and 17% used unstructured observations. (Totals add to more than 100% because participants could choose more than one format.) SDHH are taught in a variety of educational settings (Marschark, 1997). Depending on the setting, students may be educated with SDHH peers or as a single student integrated into the mainstream. The communication mode used in the classroom will also vary by setting: some schools for the deaf use sign language almost exclusively, whereas a fully mainstreamed student may have no one who signs in their learning environment (Luetke-Stahlman & Nielsen, 2003). In the “First Annual Survey,” there were some differences in alternate assessment use by educational setting. Schools for the deaf were most likely to have students participate in alternate assessments (76%), followed by district-wide/school programs (40%), and then by mainstreamed settings (12%). Findings
also varied by the percent of students with severe or profound hearing loss served by the school or program. Schools that did have students participate in the alternate assessment served an average of 58% of students with severe or profound hearing loss. In contrast, schools that did not have students participate in an alternate assessment served an average of 40% of students with a severe or profound hearing loss. There appears to be a relationship, therefore, between educational setting, student characteristics, and alternate assessment use.

Purpose of study. Previous findings demonstrate the possibility of differences in the prevalence of alternate assessment use due to both educational setting and student characteristics; however, these findings are at the school or program level and do not offer enough information to make strong conclusions about the factors that contribute to alternate assessment use. Additional information such as the state policy for alternate assessment formats and teacher recommendations for when to use alternate assessments would help to clarify factors that contribute to decisions about student participation in alternate assessment. The purpose of this article was therefore to further investigate both the prevalence of alternate assessment formats and the factors that contribute to their use.

Research questions. Three research questions guide this study:

1. What alternate assessment formats did participants report using with SDHH in 2004–2005 statewide assessments?
2. What recommendations do participants make about effective use of out-of-level testing or portfolios of student work?
3. What factors predicted the use of out-of-level testing, portfolios, and checklist alternate assessment formats in 2004–2005 statewide assessments?

Method

Instruments and Procedures

The instruments and procedures for this study were originally reported in Cawthon & the Online Research Lab (2007). Data for this paper are drawn from the “Second Annual National Survey of Assessment and Accommodations for Students Who Are Deaf or Hard of Hearing (National Survey).” The “Second Annual National Survey” was available from April through June, 2006. The survey consisted of three parts: demographics, perspectives on accommodations, and perspectives on alternate assessment. Whereas the previous publication focused on accommodations, this paper provides results from the alternate assessment component of the survey. The survey format included multiple choice, Likert scale, and open-ended response items. The survey instrument was administered in two ways: (a) online at the project Web site www.dhh-assesssurvey.org (developed using www.surveymonkey.com) and (b) paper versions provided to individuals with stamped, self-addressed envelopes for returned responses. Incentives for participation included entry in a drawing for one of 4 $25.00 gift certificates upon completion of the survey.

Participants were primarily recruited from participants in the “First National Survey” and the GRl’s Annual Survey of Schools and Programs contact list. Contacts were also made through students who are deaf and hard of hearing (SDHH) Web site affiliations, state lists of SDHH programs and services, e-mails, 687 personal postcard invitations by the principal investigator, and individual telephone calls that were made to all of the schools for the deaf in the United States. Study recruitment therefore consisted of both direct contact and “snowballing” techniques, where individuals refer their colleagues to the study through informal professional networks. Additional information about multiple recruitment strategies used to contact potential participants is available in Cawthon (2006). Most of the respondents preferred completing the online version of the survey (89%) rather than the paper version of the survey (11%). Each respondent reported on a group of students that they served or taught in the 2004–2005 school year, making the participant the primary unit of analysis for the dependent variables.

Because participants had the option of remaining anonymous (with the exception of the school or district name), it was necessary to review the data set for potential duplicate information about alternate assessment use with SDHH. After all responses were
collected, care was taken to verify whether more than one teacher reported data for the same students. Within each school or district, participants were first sorted by the grade range of the SDHH students they served. If there was any overlap in those figures, researchers then looked at data on the participation of students in alternate assessment use. If these figures were the same, the participant with the most complete set of responses, from demographic data through to best practices recommendations, was left in the data set for student and alternate assessment results. This process led to the elimination of four participants who were from the same school, grade, and had the same number of students as another participant. Data on views on validity and best practices were left in the data set.

Analysis

Open-ended coding. A key component of this survey investigated teacher recommendations of effective practices for two alternate assessment formats: out-of-level and portfolio use. We chose out-of-level and portfolio use because of their unique use with SDHH (out-of-level) or their prevalence in national policy (portfolio). We did not include a question regarding checklists due to time and space constraints on the survey instrument. The questions were open ended and allowed for participants to describe under what conditions they would recommend use of the alternate assessment format. Responses were analyzed for recurring themes using thematic content analysis. Categories reflect similar meaning of the responses, even if different vocabulary words were used to express them. Two research assistants coded each response. Both coders had developed the thematic categories and participated in analyzing pilot data. The initial interrater reliability was 93% for portfolio and 78% for out-of-level testing questions. The lead researcher evaluated responses where coders did not agree and made the final coding designation.

Because teachers often gave several examples in their answers, responses could be coded for more than one theme. Coding categories included (a) aspects of the test; (b) student characteristics; and (c) issues surrounding testing students with disabilities, such as validity, policy, student academic level, student communication, student academic level, test subject, test format, assessing progress, and additional disabilities. Examples of items for each code can be found in Appendices A and B.

Logistic regression. The third research question guiding this study centers on variables that predict the use of the three alternate assessment formats: out-of-level, portfolio, and checklists. The purpose of this analysis was to explore how different levels of the system may affect alternate assessment use. The following analyses use logistic regression with the forward method based on the Wald statistic to identify the best models for predicting the use of each format. A logistic regression is similar to a multiple regression with categorical instead of continuous dependent variables. The forward selection method was chosen because this was an exploratory research project designed to identify which variables are good predictors of alternate assessment use (Thayer, 2002). In addition, the Wald statistic was chosen because it is one of the best methods for variable selection and is more efficient than other methods (Harrell, 2001).

We examined three categories of independent variables: student characteristics, teacher perspectives, and environmental/context characteristics. These variables were drawn from previously discussed components of the “National Survey” and are summarized below.

Student characteristics. The student characteristics include candidate predictors that describe the characteristics of students served by participants in this study: (a) number of students, (b) grade level, and (c) other disabilities. Participants indicated the number of students in each grade level that they served. First, in order to create a categorical variable, the total number of DHH students served by each school program was dummy coded into one of four categories: 1–5, 6–10, 11–20, and 21 students or higher. The total number of SDHH served was dummy coded because the variable was not normally distributed. Moreover, transformations were not conducted because they are more appropriate with ungrouped data, and it is difficult to interpret the log odds of transformed variables (Tabachnick & Fidell, 2007). The categories (e.g., 1–5, 6–10, 11–20, and 21 or higher) were chosen based on the percent distribution of cases (one quartile in each group). The grade levels in the analysis included kindergarten through 12th grade (many teachers taught
students in multiple grades). If a participant served a student in a grade (e.g., fifth grade), that variable received a code of 1 = served. Finally, study respondents identified whether or not they had worked with a SDHH who had a LD, was DB, had a CD, had CP, or had an emotional disorder (ED).

Teacher perspectives. Teacher perspectives were drawn from teacher ratings of ease of use for an alternate assessment as well as their views on best practices. Ease of use ratings were shown to be significant across many accommodations used with SDHH and thus were included in our models here (Cawthon & the Online Research Lab, 2007). To facilitate interpretation in the logistic regression, the ease of using each accommodation was coded in the following way: 1 = “very difficult,” 2 = “fairly difficult,” 3 = “neither easy nor difficult,” 4 = “fairly easy,” and 5 = “very easy.” These values are “flipped” from the responses participants gave on the survey itself. In previous analysis of factors that predict assessment accommodations use, we found no differences in responses from participants who identified themselves as “teachers” vs. “administrators” in their role working with SDHH (Cawthon & the Online Research Lab, 2007). We therefore analyzed all participants and did not disaggregate this analysis by participant role.

Participant views on best practices for portfolio and out-of-level testing were drawn from the open-ended questions discussed in the previous section of this paper. For portfolio, the variables used in the logistic regression included the following reasons for recommendation: (a) a good method to see student progression, (b) good for students with disabilities, and (c) good for students who are below grade level. For out-of-level, variables included (a) whether this form of assessment helps to show student progression to parents and (b) that it should be used when student is below grade level.

Environment and context characteristics. The environment and context characteristic variables included the educational setting, communication mode in the classroom, and state policy. Educational setting variables were whether the study participant taught in a school for the deaf, regional/district program, or a mainstreamed setting. The communication modes included (a) American Sign Language (ASL), (b) other signed language, (c) oral only (speech), (d) oral and signed language together by instructor, and (e) oral by instructor and signed language by interpreter.

The policy variable was adapted from the “National Center on Educational Outcomes” report on state policies on alternate assessment (Thompson et al., 2005). Thompson et al. analyzed state assessment policies and created the following four categories: “portfolio, checklist, IEP, other, and in revision.” State policies were included for the portfolio use analysis; checklist and out-of-level testing had too many states with missing data to make it a viable predictor to include in these models. We coded all states that had a portfolio designations with $P = 1$ and the remaining states with $P = 0$. The state policy fields were merged into the National Survey database by state of residence for each participant, allowing us to summarize state policy on alternate assessment use for their students.

Dependent variables. The participation in alternate assessments section of the National Survey asked respondents about whether or not the alternate assessment was used with one or more of their students. The type of alternate assessments included checklists, portfolios, and out-of-level alternate assessment. Each of these alternate assessment variables was transformed into a dichotomous variable: none of the students served by the participant received the alternate assessment (coded as 0) or at least one student received the alternate assessment (coded as 1).

Data reduction and missing cases. The data for each model were screened for assumption violations, including multicollinearity (George & Mallery, 2006; Pallant, 2005; SPSS Training, 2001; Tabachnick & Fidell, 2007). None of the criterion variables were found to be highly correlated. For example, in generating the collinearity, diagnostics all of the standardized beta coefficients were lower than 1 and all of the tolerance levels were higher than .01, indicating that multicollinearity did not exist.

In addition to controlling for multicollinearity, the number of candidate predictor variables was reduced by using the formula discussed by Harrell (2001). The procedure identifies the number of cases required in
each group of the dependent variable in a logistic regression in order to avoid over fitting the model (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). Harnell’s formula is \( p < \frac{m}{10} \), where \( p \) is the number of candidate predictors and \( m \) is the number of cases required in each group of the dependent variable. As an illustration, in the checklist model, there are an identified 33 possible candidate predictors. Using the equation above to solve for \( m \), each group of the dependent variable (i.e., those who did and did not receive the checklist format) would require at least 330 cases. Consequently, there are not enough cases in the checklist model to use 33 predictor variables. There are only 195 respondents who did not use the format and only 97 who did use the assessment. In order to reduce the number of candidate predictors, only the predictors that were found to have a statistically significant chi-square statistic (\( \chi^2 < .05 \)) were loaded into the model (Harrell, 2001). Hence, the number of candidate predictor variables is reduced from 33 to 10. Referring back to the formula, the number of cases required in each group of the dependent variable is 100. In summary, the maximum number of candidate predictor variables that could be used was 10 for the checklist model, 12 for the portfolio model, and 7 for the out-of-level model.

We also addressed issues with missing data in this analysis. Each of the 3 alternate assessment use models had candidate predictor variables with missing data ranging from 16% to 24%. In working with missing data, the randomness of the missing data was considered to be more important rather than the amount (Tabachnick & Fidell, 2007). Variables were first examined for randomness; if randomness was found, then the missing values were randomly replaced with 0s and 1s (Harrell, 2001). If the missing data were found to be statistically significantly (\( p < .05 \)) different from nonmissing data, then the variable with the missing values was excluded from the analysis.

Results

Demographics

A total of 314 teachers or administrators from mainstreamed educational settings (\( n = 115, 37% \)), schools for the deaf or hard of hearing (\( n = 80, 25% \)), or district and regional programs (\( n = 119, 38% \)) participated in this study. Respondents served students as teachers of the deaf (50%); regular education, special education, or itinerant teachers (20%); administrators (8%); and those serving in multiple roles (6%). Participants lived in all regions of the country with at least one participant in each state (including the District of Columbia). We do not have estimates of the national population of teachers and educational professionals who served SDHH, but we do have estimates of the student population. Although the sample was not random, the results are still fairly representative of SDHH throughout the United States (Table 1). As an example, 11% of the students in the National Survey were from the northeast whereas an estimated 17% of SDHH nationwide live in that Census region (Mitchell, 2004). The south is overrepresented, with 41% of this sample compared to a population estimate of 32%. The proportion of students in schools for the deaf (53%) far exceeds the national average, an estimated 25% GRI (2005). Student characteristics, school resources, and instructional strategies may vary greatly from region to region; greater representation from the northeast and from mainstreamed settings is thus necessary to make stronger conclusions about how each factor contributes to alternate assessment use.

Language used in instruction. The survey collected information about the primary communication mode used in instruction. Participants indicated the most prevalent mode of communication used in instruction; this does not account for the likely variability in communication used with specific students in each classroom or school. Data for the most common forms reported are shown in Table 2: ASL, other signed language (including Pidgin Sign and Signed Exact English), oral (speech) only, total communication (oral and sign together by instructor), oral (speech) plus interpreter, and other. Results show that a range of communication modes was used in all settings. Participants in schools for the deaf were more likely to use ASL (88%), those in district/regional programs used oral and sign language together (79%), and those in mainstreamed programs used either oral only (67%) more than other communication modes.
Hearing loss and cochlear implants. Deaf or hard of hearing students experience a wide range of hearing impairment, from mild to profound hearing loss. In this study, participants indicated the level of hearing loss of the students they served. We then calculated the mean percent of students at each level for the students served by study participants. Mainstreamed settings had a higher mean percentage of SDHH with a mild hearing loss ($M = 21\%$) than schools for the deaf ($M = 6\%$) or district or regional programs ($M = 12\%$), $F(2, 243) = 11.55, p < .001, \eta^2 = .09$. In contrast, schools for the deaf had a higher mean percentage of students with profound hearing loss ($M = 49\%$) than either district or regional programs ($M = 29\%$) or mainstreamed settings ($M = 18\%$), $F(2, 244) = 18.361, p < .001, \eta^2 = .18$. We also considered the number of students fitted with a cochlear implant. In this sample, students with cochlear implants are evenly represented within schools for the deaf ($M = 10\%$) and mainstreamed settings ($M = 13\%$), but higher in district/regional programs ($M = 20\%$), $F(2, 313) = 5.43, p < .01, \eta^2 = .03$.

Additional disabilities. Finally, participants described the prevalence of additional disabilities in the SDHH they served. Nearly 80% of study participants indicated that they served at least one student with a LD. Other frequently noted disabilities included CD (51%), ED (35%), CP (26%), deaf-blindness (DB) (22%), or another disability such as attention deficit and hyperactivity disorder (35%). Only 12% of participants indicated that none of their students had an additional disability.

Alternate Assessment Use

Alternate assessment formats and strategies investigated in this study were based on the National Center for Education Statistics (2001) and Mitchell (2004).
on Educational Outcomes list of common formats used in state testing (Thompson & Thurlow, 2003) and from the results of the first National Survey (Cawthon & the Online Research Lab, 2006). Participants noted which of the following three alternate assessment formats were used with at least one of their students: checklist, portfolio, and out-of-level assessments (Table 3). Percentages represent the proportion of teachers from each setting who reported whether or not they used the format. Totals may add to more than 100% because participants could choose more than one format if the students they served used multiple forms of alternate assessment. The number of participants responding for each format was \( n = 284 \) for checklist, \( n = 294 \) for portfolio assessments, and \( n = 254 \) for out-of-level. A total of \( n = 92 \) (32%) participants said that they used a checklist with their students, \( n = 111 \) (40%) used a student portfolio, and \( n = 57 \) (23%) used out-of-level testing with their students (participants could choose more than one format). Educational setting was an important factor in the likelihood that students participated in alternate assessments. Teachers from schools for the deaf were significantly more likely than teachers in district programs or mainstreamed settings to use each of these alternate assessment formats (checklist, \( \chi^2(2, N = 284) = 17.18, p < .001, \chi^2 = .246 \); portfolio, \( \chi^2(2, N = 281) = 21.16, p < .001, \chi^2 = .274 \), and out-of-level, \( \chi^2(2, N = 244) = 10.8, p < .01, \chi^2 = .21 \)). These results confirm school-level findings from the First National Survey.

### Table 3  Use of alternate assessment formats in schools for the Deaf, district/regional programs, and mainstreamed settings in the 2004–2005 state standardized assessments

<table>
<thead>
<tr>
<th>Alternate assessment format</th>
<th>Educational setting</th>
<th>Schools for the Deaf</th>
<th>District/regional programs</th>
<th>Mainstreamed settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Checklist</td>
<td>37 (52%)</td>
<td>31 (28%)</td>
<td>24 (24%)</td>
<td></td>
</tr>
<tr>
<td>Portfolio</td>
<td>43 (62%)</td>
<td>43 (28%)</td>
<td>37 (36%)</td>
<td></td>
</tr>
<tr>
<td>Out-of-level</td>
<td>21 (40%)</td>
<td>20 (20%)</td>
<td>16 (18%)</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Number and percentage of respondents who used the alternate assessment format with their SDHH for the 2004–2005 state assessments used for NCLB accountability purposes.*

Participant Ratings on Ease of Use

Participants gave their perspectives regarding the ease of use for each of the three alternate assessments investigated in this survey (Figure 1). Participants rated accommodations on a scale of 1–5, with lower mean scores associated with an easier to use accommodation. Overall, the checklist (\( M = 1.88, SD = .99 \)), portfolio (\( M = 2.39, SD = 1.2 \)), and out-of-level (\( M = 2.48, SD = 1.23 \)) assessments were rated as easy to implement. None of the assessment formats were seen as difficult to implement (with a mean score of 3.0 or greater).

We performed a one-way between-groups analysis of variance to investigate the perceived ease of use for each accommodation in the three different educational settings. There was one statistically significant difference at the \( p < .01 \) level in the analysis: participants from mainstreamed settings rated the checklist format as more difficult than those in district/or regional programs (\( F(2,212) = 5.8, p = .003, \eta^2 = .05 \)). Post hoc comparisons using the Tukey HSD test indicated that the mean score for mainstreamed settings (\( M = 2.16, SD = 1.10 \)) was higher from that of the district/regional programs (\( M = 1.64, SD = .87 \)). Schools for the deaf (\( M = 1.94, SD = .92 \)) did not differ significantly from either mainstream settings or district/regional settings. There were no statistically significant differences between educational settings in perceived ease of use for portfolio and out-of-level alternate assessment formats.

 Perspectives on Alternate Assessments

Some teachers were advocates for using alternate assessments, in general, with SDHH. However, when asked about factors that are important to consider when using a specific format, responses were different for portfolios than for out-of-level testing. A summary of teacher perspectives for both formats is given below.

**Portfolio.** An overview of participant suggestions for best practices in portfolio use is shown in Figure 2. Participant responses were coded across a list of 14 categories, plus one category for “other.” The most prevalent recommendation for portfolio use was because it was a good tool to assess progress, particularly for students with additional disabilities. For examples
of participant responses to the portfolio question for each category, please see Appendix A.

Participants gave a wide range of responses to this survey question. The most frequently cited reason for using a portfolio was as a means for measuring student progress or change in performance over the school year (27%). A similar percentage of participants felt that the portfolio was not applicable for their students, either due to student characteristics or availability in their state (24%). Nearly a quarter of responses were unique and uncodable with responses from peers, resulting in a large proportion of answers in the “other” category (23%). This may reflect the great diversity of students served by study participants as well as the large

---

**Figure 1** Mean scores for ease of use for alternate assessments respondents in schools for the Deaf, district/regional programs, and mainstreamed settings in the 2004–2005 state standardized assessments. Data represent teacher ratings on ease of use for each alternate assessment format. Ratings were on a 5-point, Likert scale (1–5), with lower scores indicating an easier to use alternate assessment format. Participants could also indicate that they had no opinion; these scores were not included in this analysis.

**Figure 2** Perspectives on portfolio use for SDHH. Open-ended responses to questions about their perspectives about portfolio use for SDHH. Figures represent the percent of all participants whose responses were coded for each category. If participants gave multiple recommendations, their response may have received more than one code. If the participant said that the format was not applicable or available to their students, the response received the code “Not Applicable.”
range of issues that arise when discussing assessment practices. The remaining categories had response from 2% to 16% of participants, drawing on issues such as student characteristics (e.g., additional disabilities or communication mode), test characteristics (e.g., subject of the test), and communication with parents about student academic progress.

Out-of-level. Figure 3 provides a similar overview of participant recommendations for best practices in out-of-level testing. Participant responses were concentrated into fewer categories than portfolio recommendations. Responses to the question about out-of-level testing were more strongly skewed either supporting its use or with concerns about the implications of testing students below grade level. For examples of participant responses to out-of-level for each category, please see Appendix B.

The range of responses to out-of-level testing was less broad than for the portfolio question. Responses were coded across 10 categories. Most of these categories overlapped with those found in the portfolio question analysis. However, limited use was a category unique to out-of-level testing as an alternate assessment format for SDHH. Two percent of the respondents indicated that out-of-level testing should have or only has limited use with their students. In a related vein, participants noted that out-of-level is not applicable to their students (28%). Those who did offer perspectives on when it might be used noted that it would be appropriate when there were concerns about the student’s academic level being too far behind their peers for meaningful participation in grade-level assessments (22%) or when it would help in measuring student progress (15%).

Predictors of Alternate Assessment Use

The following results are from direct logistic regressions of predictors of alternate assessment use with SDHH. This analysis produces an odds ratio to determine the strength of the effect of predictors on outcomes (Tabachnick & Fidell, 2007). The odds ratio is a way of comparing whether the probability of a certain event is the same for two outcomes of a single predictor. In these results, odds ratios show the relatively likelihood of using a format when the predictor is present (such as a state policy that allows the format) vs. when it is absent. An odds ratio of 1.0 indicates equal likelihood of occurring in either outcome. As the odds ratio becomes larger than 1.0, the outcome is more likely to occur in one condition than in the other.

Portfolio alternate assessment. Descriptive information for each predictor of portfolio use is shown in Table 4. Table 5 shows results for the portfolio format. A test of the full model vs. a model with intercept only was statistically significant, \( \chi^2(5, N = 285) = 47.9, p < .001 \). The Hosmer–Lemeshow test yielded a \( \chi^2(8) \) of 6.349 and was nonsignificant (\( p = .608 \)). As the odds ratio value increases, so does the impact of that predictor variable on alternate assessment use. Results indicate that the odds of portfolio use are approximately three times higher at a school for the deaf (odds ratio = 2.94), over two times higher when the portfolio is believed to be a good tool to show student progression (odds ratio = 2.67), and almost twice as likely when the participant served students in kindergarten, 7th, 8th, 11th, and the 12th grades (odds ratios of 1.92
for kindergarten, 1.83 for both 7th, 8th, and 1.81 for both 11th and 12th).

**Checklist alternate assessment.** Table 6 provides descriptive data for each predictor of checklist alternate assessment format.

Checklist alternate assessment. Table 6 provides descriptive data for each predictor of checklist alternate assessment format.

Table 5  Direct logistic regression summary for portfolio

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Beta coefficient</th>
<th>Wald $\chi^2$</th>
<th>Odds ratio</th>
<th>95% CI for odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergarten</td>
<td>.650</td>
<td>5.64*</td>
<td>1.92</td>
<td>1.12</td>
</tr>
<tr>
<td>Grade 7 or 8</td>
<td>.607</td>
<td>4.92*</td>
<td>1.83</td>
<td>1.07</td>
</tr>
<tr>
<td>Grade 11 or 12</td>
<td>.593</td>
<td>4.50*</td>
<td>1.81</td>
<td>1.05</td>
</tr>
<tr>
<td>School for the Deaf</td>
<td>1.078</td>
<td>12.14***</td>
<td>2.94</td>
<td>1.60</td>
</tr>
<tr>
<td>Good tool to see student progression</td>
<td>.983</td>
<td>10.44***</td>
<td>2.67</td>
<td>1.47</td>
</tr>
<tr>
<td>Constant</td>
<td>−1.726</td>
<td>38.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$p < .05$. *** $p < .001$. 

Table 7 shows the results for the checklist format. A test of the full model vs. a model with intercept only was statistically significant, $\chi^2(2, N = 275) = 19.008, p < .001$. The Hosmer–Lemeshow test yielded a $\chi^2(2)$ of .188 and was nonsignificant ($p = .910$), suggesting that the model fits the data well. Results indicate that the odds of checklist use were almost three times as likely at a school for the deaf (odds ratio $= 2.73$) and almost twice as likely when the participant served students in grade 11 (odds ratio $= 1.81$). The other predictor variables did not show significant odds ratios for checklist alternate assessments.

**Out-of-level alternate assessment.** Table 8 provides the descriptive data for out-of-level predictors.

Table 9 shows the results for out-of-level assessment. A test of the full model vs. a model with intercept only was statistically significant, $\chi^2(4, N = 247) = 30.434, p < .001$. The Hosmer–Lemeshow test yielded a $\chi^2(6)$ of 2.818 and was insignificant ($p = .910$). There were several predictors that were significant in this analysis. Results indicated that the odds of out-of-level use was three times higher if the respondent thought out-of-level was best for when students were below grade level (odds ratio $= 3.00$), twice as likely if the instructional communication was in ASL (odds ratio $= 2.02$), twice as likely if the instructional communication was in another form of signed language (odds ratio $= 2.34$), and almost twice as likely if the participant served 11th grade students (odds ratio $= 1.86$).

**Discussion**

The purpose of this article was to present findings on the use of alternate assessments with SDHH. We reported findings from the “Second Annual Survey of
Assessments and Accommodations for Students Who Are Deaf or Hard of Hearing,” a national survey with participants who serve students in a variety of educational settings. The overall findings regarding alternate assessment format use built on those from the First Annual Survey (Cawthon & the Online Research Lab, 2006). Whereas the previous survey focused on alternate assessment prevalence in the school or program, the current analysis reported findings at the teacher level. Although not as fine grained as individual student data, the teacher (or participant) level narrowed the findings down to the alternate assessment trends ranging from a single classroom (at schools for the deaf) to a single student (for a student mainstreamed alone). All three formats researched in this study were used by at least a quarter of study participants, indicating that alternate assessments formats are prevalent in testing SDHH.

This discussion will focus on some key issues in this analysis: limitations to the study design, the role of a portfolio format, role of checklists, the role of out-of-level testing, the significance of educational setting, and implications.

Limitations

There are a number of limitations to this study design that affect the strength of conclusions that can be drawn from these results. The first is true of any survey design that allows participants to opt for an anonymous participation. It is a challenge to confirm the accuracy of the self-report data when some individuals cannot be contacted for clarification. This was a criterion set forth by the institution’s Internal Review Board in anticipation of the potentially sensitive nature of the subject. A second limitation of the study is...
the time lag between the time students took the assessments and when the participants report on assessment use. Although each participant, particularly teachers in mainstream settings, have relatively few SDHH in their caseload or classroom, there may be some inaccuracies due to the time delay. Because each state has standardized testing at different intervals, a national study necessarily has to occur after the end of previous school year to capture all states. Summer vacations push the start date for the survey such that it may occur as much as a year after the assessment period under investigation (for those who test in the fall of each academic year). Third, there is still a relatively high level of missing data for some questions on the survey. Although this has improved over previous surveys and data reduction strategies partially address the potential bias in missing data, the study would be strengthened with more complete participant responses. Finally, this survey did not ask questions about how validity issues are addressed at the local level. For example, participants were not provided the opportunity to describe who administers or evaluates the assessment or how individual scores were used. The impact of alternate assessment on individual scores and their inclusion in NCLB reports is an area that needs further research.

Role of Portfolio Format

Portfolios were the most common format of an alternate assessment used by participants in this study. Nearly two-thirds (65%) of participants from schools for the deaf used a portfolio with at least one of their students, roughly twice the proportion of participants from district programs (32%) and mainstreamed settings (37%). Participants rated portfolios as easy to use, results that stand in contrast to discussion regarding the challenges of alternate assessment use. From a policy perspective, portfolios are one of the most commonly allowed formats, currently in use as the format for half of the states (Thompson et al., 2005). Understanding what role teachers see portfolios playing in the assessment of SDHH is therefore a timely and relevant question. These recommendations for best practices highlight the importance of (a) how the portfolio is compiled, (b) who participates using a portfolio, and (c) how the scores are used to monitor progress.

The perspectives from participants provide us with insightful thoughts about how portfolios may be a strategic tool for including SDHH in accountability assessment frameworks. The most frequent comment about portfolios is their ability to help teachers monitor growth over time. Because the portfolio allows for work samples throughout the year, evaluators can see how a student’s mastery or knowledge and skills changes over time. This finding is in contrast to summative, end-of-year information from statewide, standardized assessments. Scores from these assessments do not show teachers and parents the specific areas of progress or delay in an individual student’s learning. The structure of portfolios may therefore provide more information to teachers and parents than is currently available in many standardized assessment systems.

A second theme in participants’ responses focused on student characteristics: the use of portfolios for students with multiple disabilities or those who are significantly below grade level. Teachers who responded in this category focused on disabilities that prevented a student from participating in a standardized assessment, even with accommodations. The kinds of disabilities mentioned in these responses included those with cognitive delays or with autism. It may be that these responses are due to the fact that many state policies limit the use of alternate assessments to students with significant CDs. Although the

Table 9  Direct logistic regression summary for out-of-level

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Beta coefficient</th>
<th>Wald $\chi^2$</th>
<th>Odds ratio</th>
<th>95% CI for odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Served Grade 11</td>
<td>.62</td>
<td>3.98*</td>
<td>1.86</td>
<td>1.01 3.42</td>
</tr>
<tr>
<td>Instructional communication was ASL only</td>
<td>.70</td>
<td>4.86*</td>
<td>2.02</td>
<td>1.08 3.78</td>
</tr>
<tr>
<td>Instructional communication was other signed language</td>
<td>.85</td>
<td>5.87*</td>
<td>2.34</td>
<td>1.18 4.64</td>
</tr>
<tr>
<td>Student is below grade level</td>
<td>1.10</td>
<td>11.24***</td>
<td>3.00</td>
<td>1.58 5.71</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.23</td>
<td>52.90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.  ***p < .001.
number of grades below their peers was not always quantified, teachers often used descriptors such as “so low,” “significantly delayed,” and “not able to meet the benchmarks for the curriculum.” Regardless of the reason for the delay, be it due to hearing loss, language skills, or additional disabilities, these participants focused on how portfolios can be used with students who are working on below grade-level academic skills.

A final theme in the responses about portfolios focused on the role of the teacher in collecting classroom artifacts for the assessment. Several teachers remarked that the alternate assessment was quite labor intensive and required diligent data collection by the teacher throughout the school year. States vary in what structure is given to both how the portfolio is compiled and how the evaluations are made. One participant remarked on the importance of not only regularly collecting artifacts but also being familiar with how the portfolio is scored in order for the materials to fairly represent student progress and achievement. Finally, another participant noted the need for a comparison set of artifacts from peers so that the students’ body of work can be compared with those who completed similar tasks. These perspectives serve as a timely reminder that the accuracy of a portfolio depends on adequate time and training in how to compile them, a particularly important aspect of using portfolios within accountability reforms.

Role of Checklists

Although the information regarding checklists is more limited than for the other two formats (we do not have open-ended responses for this format), there are a few findings from this analysis that help our thinking about the role of checklists in assessing student performance in an accountability framework. Checklists of student knowledge and skill are used by a sizable proportion of participants in this study, ranging from over half of the teachers from schools for the deaf and a quarter of participants from the regional programs or mainstreamed settings. This is in contrast to the small number of states ($n = 7$) that allowed checklists as the alternate assessment format for accountability purposes for the year these data were collected. Participants also rated checklists as easy to use, indicating that difficulty in completing a checklist should not be a barrier to use. There were only two factors that significantly predicted different rates of checklist use: school for the deaf and serving students in 11th grade. More participants from schools for the deaf reported using alternate assessments of all formats, and checklists are included in that trend. The descriptive statistics show a greater tendency for participants with students in high school to use a checklist of skills (only 11th grade was a significant predictor in the logistic regression). The reasons for this are unclear, and an area for further exploration. Speculatively, it might be that checklists are easier for teachers to complete when they teach in multiple classrooms across the school day. In contrast with elementary school teachers who work with a small group for the entire day, secondary level teachers often teach across grades and a larger number of students. It may also be the case that the skills measured when children are younger are more amenable to portfolio assessments because they are formative in nature, whereas topics such as advanced math and science content can be measured with a shorter, end-of-term evaluation. However, because portfolio use was high overall, including in the upper grades, it is unclear what underlying factors contribute to the higher use of checklist in these grades.

There were a number of characteristics that did not predict checklist use in this logistic regression model. Not on this list include the type of language used in instruction (e.g., ASL, oral only, sign language interpreter), students having additional disabilities, or number of students served by the respondent. Of these, languages of instruction stands out as different than the predictors for out-of-level assessment practices (see discussion below). Checklist use does not appear to be sensitive to language differences in how teachers provide instruction in the classroom. This result is reasonable considering that checklist process does not require evidence of student knowledge that is communicated either via visual or written form. Instead, teachers provide an evaluation based on student performance within classroom activities or steps toward IEP goals. In this way, perhaps checklists provide potential for more equitable access to the content of the assessment than formats that require submission of student products or participation in a standardized assessment.
Role of Out-of-Level Testing

Testing out-of-level is a significant issue because many deaf or hard of hearing students are below grade level in reading and mathematics (Traxler, 2000). Out-of-level testing has played a unique role in assessment for SDHH. Because a form of out-of-level testing has been used as part of the development of DHH norms for the Stanford achievement test, practitioners have used out-of-level data as part of the assessment process for their students. The prevalence of out-of-level testing found in the current analysis may be partially due to this historical precedent. Participants also rated them as easy to use, not surprising given the standardized nature of this format. Nearly a quarter of participants, 22% overall, reported using an out-of-level test with at least one of their students, with a larger proportion of participants from schools for the deaf reporting their use (41%) than those from district programs (25%) or mainstreamed settings (19%). Although not the most predominant form of alternate assessment, it is still used to a larger extent than would be expected given the restricted nature of state policy.

Participants who supported the use of out-of-level testing indicated that they felt it to be a helpful tool in measuring progress, with one participant suggesting that the change in performance to be a way of showing closures in the achievement gap with hearing peers over the course of a school year. These suggestions were often qualified with the concern that the scores not be aggregated with on-level standardized test scores but could be used in an overall measure of student learning. Other statements of support focused on the fit between the reading level of the student and the grade level of the assessment. If the student is reading significantly below grade level, some participants felt that out-of-level assessments would be more appropriate than using a grade-level test. This perspective questions the utility of a test that is too far above the reading level of the students to yield meaningful results. The concerns about student reading level parallels responses that participants gave in support of accommodations that allow administrators to present test items in ASL instead of having the student read the test item (Cawthon & the Online Research Lab, 2007). The out-of-level testing option is seen as one potential strategy for creating a test that clearly demonstrates a student’s proficiency level while at the same time reducing frustration for students faced with test items that they do not understand. This premise for using alternate assessments is therefore different than the guidelines put forth by NCLB. NCLB designates alternate assessments as appropriate for those with the most severe disabilities. It is not the severity of the disability, per se, that guides these teachers’ use of alternate assessment but the match between how to measure student progress and the format of assessments.

Not all survey participants supported the use of out-of-level testing for SDHH. A number responded vehemently that out-of-level testing should never be used as part of the statewide assessments. Concerns arose about how the out-of-level tests were administered, how the scores could not be used in accountability because they cannot be aggregated with other scores (and thus were seen as a “waste of time”), and that their state policies strongly discouraged or forbid their use. Indeed, out-of-level test scores are limited in their utility in a grade-level standards-based accountability system. Interestingly, state policies did not significantly predict use of out-of-level testing in the regression analysis. It was unclear from some responses whether a state policy that allowed out-of-level testing would alleviate concerns about its use. For example, one participant stated that “they should be taught as the same level as their peers and should be tested accordingly. Anything less is a failure of the school system.” This illustrates a deeper, philosophical objection to evaluating SDHH using different standards than their hearing peers.

Educational Setting

Educational setting, in this article, is defined along three main categories: schools for the deaf, regional/district programs, and mainstreamed settings. It should be noted that there can often be overlap between these categories and that some caution is warranted in drawing strong conclusions about differences in findings between educational settings. This said, there are recurring themes in the results that point to the importance of educational setting in factors predicting the use of alternate assessments.
Participants in schools for the deaf use alternate assessments to a greater extent than participants in district programs or regional settings. This is perhaps not surprising for a number of reasons, including (a) teachers in schools for the deaf tend to serve a greater number of students, thus increasing the likelihood that one of them may use an alternate assessment and (b) students who attend schools for the deaf are more likely to have profound to severe hearing loss, with greater ramifications for literacy development. A number of the significant predictors found in the exploratory logistic regression fit this trend: working at a school for the deaf (checklist and portfolio), serving students in a high school grade (checklist, portfolio, and out-of-level), and using ASL in instruction (out-of-level). These variables all support the finding that whether viewing assessment practices at a school or at a teacher level, students who attend a school for the deaf are more likely to use a portfolio for assessment than students in other educational settings. Student-level analyses of alternate assessment use would help to further clarify the individual roles of student characteristics, teacher perspectives, educational setting on these outcomes.

For example, there were two factors predicting alternate assessment use that may not be directly tied to educational setting. These factors arise from the best practices recommendations offered by participants for portfolio use and out-of-level testing. When teachers recommended portfolios because they were a good tool to use to see student progress, at least one of their students was over 2.5 times more likely to receive the portfolio format than when they did not. Similarly, when teachers recommended out-of-level testing because a student’s academic knowledge and skill are below grade level, at least one of their students was three times more likely to receive out-of-level assessments. Although only preliminary, these findings point toward “teacher perspectives” as an important factor, along with student characteristics and educational setting, in predicting alternate assessment use.

Implications

Taken together, results from this analysis can provide some guidance to professionals who make assessment decisions for SDHH, particularly in the context of NCLB accountability frameworks. One strong criticism of accountability frameworks is that assessments do not provide sufficient information to guide future instruction (Cawthon & the Online Research Lab, 2006, 2007). A central question raised by teachers in this study includes: Do standardized tests that are several years above student academic level still yield valid scores? Although a score of “not proficient” on the grade-level standardized assessment is certainly accurate, the assessment process may not provide meaningful information for students, parents, and teachers to use in improving educational outcomes. This is because the test will show very little of what the student does know. A segment of this study sample, particularly teachers with students who are not proficient readers, advocated for some form of alternate assessment to be available to measure student learning. Alternate assessments, though fraught with their own validity concerns, will continue to be proposed as a valuable component to an accountability assessment framework.

A related finding from this study is that portfolios may best fulfill two needs from assessment: (a) to provide a measure of school effectiveness for accountability and (b) to understand the progress students make toward IEP goals. If and when states have clear guidelines on how teachers are to assemble and score alternate assessments, it is possible that portfolio assessments can fill the needed gap between the goals of state assessment systems and those of teachers who serve SDHH. Further work on ensuring the reliability and validity of portfolios will be essential in providing access to fair alternate assessments.

A third implication of these results is the need for caution in using out-of-level assessments for accountability purposes. Whereas there is a history of out-of-level testing in the development of SAT norms for SDHH, those procedures do not translate well within standardized assessment practices under NCLB or other measure of school effectiveness. The goal of accountability reform is to support grade-level instruction and assessment for students who have previously been excluded from this process. Although on-level standardized assessments may not be appropriate for all students, other avenues for meaningful alternate assessment may be more closely aligned to the goal of providing grade-level instruction to all students than out-of-level testing.
Setting continues to be an important factor in understanding the prevalence of alternate assessments for SDHH. This study confirmed previous findings that students at schools for the deaf receive alternate assessments more often than those in other settings (Cawthon & the Online Research Lab, 2007). However, there are many differences in the characteristics of schools for the deaf population over those in district programs or mainstreamed settings. Findings extend previous research by showing that differences by education setting may be more closely tied to age of students (those in secondary grades) or language used in instruction (some form of sign language) than student characteristics such as having an additional disability. In addition, there does not appear to be a “critical mass” factor such that settings with a greater number of SDHH are more likely to administer alternate assessments.

There are a number of unmeasured, underlying characteristics that may lead to greater prevalence of alternate assessments in schools for the deaf. For example, teachers at schools for the deaf tend to serve a greater number of students and thus may be more comfortable with the alternate assessment process than those working with fewer students. Although differences in number of students did not significantly predict alternate assessment use, this variable was transformed into a categorical, rather than continuous variable, potentially masking the effects. The status of schools for the deaf as a “host” school for student scores is another potential factor in assessment choices. From an accountability standpoint, only half of publicly funded schools for the deaf receive their own report cards (Cawthon, 2006). Most of the remaining schools either have student scores sent to the “referring” or sending district or have them aggregated to the district or state level. Student participation in alternate assessment therefore only affects Adequate Yearly Progress ratings for a limited number of schools for the deaf. This variation in accountability policies may have an indirect effect on the assessment practices in different settings.

These findings have implications for other students who may be performing significantly below grade level and yet not belong to the group of students who are allowed to take a statewide alternate assessment and have their proficiency scores counted toward adequate yearly progress. This study emphasizes the importance of the current discussion of alternate standards and who is eligible to be measured along these criteria. As discussed in the literature review, the content of standards remains the same but the depth of questions or benchmarks may be changed to meet the needs of students with disabilities. These guidelines answer, in part, concerns by scholars that alternate assessments are used only to demonstrate below grade-level proficiency in academic areas (Weiner, 2006). Teachers and educators may also find modified standards to be more appropriate for their students, particularly if they can provide feedback during the development of modified standards and implementing ways to measure and interpret student scores on these tests. In addition, as states move toward growth models that show change in student proficiency from year to year, questions of scoring, measuring progress over time, and how alternate assessment scores are included in discussions of school efficacy will only grow in their importance. Both conceptual and technical issues about the purpose of alternate assessment in large-scale accountability will hopefully gain greater clarity as state assessment systems evolve.

Conclusion

NCLB has placed issues of student achievement, measurement, and inclusive assessment systems into the forefront of school accountability. Questions of whether students participate in accountability reform have largely been answered: inclusion in assessments used to evaluate school effectiveness is expected for all students with disabilities. States have responded to this expectation with increasing clarity in their policies on how students with disabilities participate in large-scale assessments, including guidelines on accommodations use and alternate assessment formats. Yet alternate assessments represent a very different “method” of evaluating student knowledge and skill than the standardized assessments used for the majority of students. When aligned to grade-level standards, performance-based measures such as a portfolio or checklist gives teachers potentially rich ways of evaluating a student’s progress on grade-level content. It is therefore critical to dig deep and look closely at the
role alternate assessments play in how SDHH participate in statewide assessment systems. In the end, the strength of alternate assessment use, particularly when viewed from both a conceptual and technical perspective, may be a significant benefit to students, teachers, and schools seeking to improve student achievement.

**Funding**

Walden University Faculty Excellence Fund.

**Appendix A**

**Table A** Example participant responses from open-ended questions on portfolio use

<table>
<thead>
<tr>
<th>Category response type</th>
<th>Participant responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessing progress</td>
<td>When my students who were deaf also had additional disabilities that interfered with their ability to participate in an assessment or assessment with accommodations. Very helpful for all students, this is a good assessment that includes student work at various points throughout the year and includes their best work samples. I feel that a student portfolio is an accurate way to demonstrate a student’s academic growth. I feel that teachers have to have a clear set of guidelines to follow in selecting student work for a portfolio. Some students cannot test well—this is to their reading- and writing-level portfolios demonstrate that the student understand the concept or has mastered a skilled using other forms of expression rather than reading and writing.</td>
</tr>
<tr>
<td>Additional disabilities</td>
<td>These key proof that a student has learned and progressed during the school year I think these portfolios are only used for the lower IQ lower functioning life skills students at our school. Students with moderate to severe CD or autism. When my students who were deaf also had additional disabilities that interfered with their ability to participate in an assessment or assessment with accommodations. This is utilized for cognitively delayed students whose English language skills are limited.</td>
</tr>
<tr>
<td>Academic level</td>
<td>Yes this is appropriate when the reading level and academic level are significantly delayed. When the student is so low academically that this is the only way to show progress. When the student cannot take a test at his or her reading math level. Any deaf child working below grade level. They should be graded on the same level as their IEP is taught on. If the child is working on 2nd grade skills then they should not be given a sixth grade test. It is easy to see they cannot pass the test and they learn to just bubble in the answers. You cannot compare that method with a child who is truly able to take the test. When results of standardized language instruments are at the percentile or below. The student will not be able to meet benchmarks of the curriculum.</td>
</tr>
<tr>
<td>Other</td>
<td>Some students are just not good test takers. I think this is appropriate for all students. In our state, this alternate assessment is far too labor intensive to utilize unless the student is unable to function in any capacity. Very appropriate. Warn the teachers that they need to collect work samples, and work samples from peers at the beginning of the year.</td>
</tr>
</tbody>
</table>

**Note**

1. There are situations where these 3 settings may overlap, such as when a school for the deaf provides services to students in mainstreamed settings. Participants selected which of the 3 settings best described their setting. The question description included the following information: “school for the deaf” (residential or day program in a separate setting); “district-wide or regional program for the deaf or hard of hearing” (students are in special programs for part or all of the day); “mainstreamed setting” (students are in regular education classrooms for all of the day, may include itinerant or interpreter services); and “other” (please specify).
Appendix B

Table B  Example participant responses from open-ended questions on out-of-level testing

<table>
<thead>
<tr>
<th>Category</th>
<th>Participant responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not appropriate</td>
<td>This can be used when you have students that can read at least at the first grade level. Because my students are basically nonreaders, this is not a good way to assess their abilities. Our state never figured out a good way to provide this kind of assessment. It never was very successful. They tried an online off-grade level assessment but it never was straightened out. Never, because out-of-level testing procedures do not give the student a comparison of his grade-level peers and he/she can never have a proficient score. I am assuming this means giving a test based on student's age rather than functioning level. I think this is a waste of precious time. Our self-contained students must take the annual standardized test based on age. The process and results are discouraging to student, teachers, and parents.</td>
</tr>
<tr>
<td>Assessing progress</td>
<td>For a 3-year evaluation to determine level of performance is an appropriate use. It can be helpful for assessing individual growth but normed results should not be interpreted in the standard fashion. This assessment can also provide a measurable rate of growth in one or several years. The measurement of their growth against a “normal” student population while usually lower does present a picture of some gap closures during the academic year.</td>
</tr>
<tr>
<td>Academic level</td>
<td>About a fifth of participants focused on the use of out-of-level in assessing the skills students have at a level appropriate to their literacy level. For example, one participant noted: “I would recommend this accommodation when a student is being taught on a lower level and would achieve at the lower level.” Those who were opposed to its use were quite clear in their stance: “There is no off-level grade testing in this state. The majority of us would tell you that testing on grade level for many of our students is a farce.” I would never recommend this for a student whose only disability was a hearing loss. They should be taught at the same level as their peers and should be tested accordingly. Anything less is a failure of the school system. We cannot administer a state test at a varying level other than the grade level the student is placed within. We can, however, administer a different level of district testing—we consider the child's instructional level, communication mode/ease, as well as written language and reading abilities.</td>
</tr>
<tr>
<td>Policy</td>
<td>We have state guidelines that outline the students who are appropriate to take an out-of-level administration. Generally, these are students with low cognitive abilities. The state does not accept out-of-level testing as valid. In the state, we are not allowed to even consider this option. I'm not sure if this is an appropriate testing. The state has developed standards for alternate testing, and they seem to be appropriate for students with CDs. Through this assessment, students graduate with special diploma. In California, if a student takes the standardized test, not the California Alternate Performance Assessment, out-of-level, then the student is automatically graded “BELOW BASIC.” So we are told not to give the test out-of-level. Our students usually score poorly on the standardized test, the STAR. In my opinion an “out-of-level administration of a standardized test” would be most appropriate for my students who are not reading on grade level. I do not understand why we are forced to administer a grade level test to a child who reads/writes well below his/her grade level. In PA, we are not permitted to administer an “out-of-level administration standardized test.” Children with IEP's (except for mental retardation diagnosis) must take the Pennsylvania state assessment. Students with a diagnosis of mental retardation take the Pennsylvania alternate state assessment.</td>
</tr>
</tbody>
</table>
References


Received January 24, 2008; revisions received May 27, 2008; accepted June 3, 2008.