

FOCUS ON EXCEPTIONAL children

Equity for Students With High-Incidence Disabilities in Statewide Assessments: A Technology-Based Solution

Ed Meyen, John Poggio, Soonhwa Seok, and Sean Smith

One of the most significant challenges facing policy makers in education today is to ensure that state assessments designed to measure student performance across specified grade-level curriculum content standards will allow all students to demonstrate what they have learned. This challenge is made complex by the varied attributes of students with disabilities and the curriculum these students receive.

The scope of the complexity became particularly evident with passage of No Child Left Behind (NCLB), the 2001 revision of the Elementary and Secondary Education Act (ESEA), which mandates that schools, districts, and state departments of education be held directly accountable for the progress of all students, including students with disabilities (Allbritten, Mainzer, & Ziegler, 2004). Thus, school districts must demonstrate that students are making ongoing progress toward proficiency, and that all students reach the designated proficiency level by the year 2014. Attendance centers, districts, and states that fail to meet the stated requirements must provide supplemental services. Further, continued failure to make progress toward proficiency will result in drastic changes and sanctions.

According to the 24th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act (IDEA) of 2002, more than 2.5 million children ages 6–17 with disabilities were served under IDEA in the 50 states and the District of Columbia. This represented 11.05% of the school population (U.S. Department of Education, 2002).

Given the stringent accountability mandates and the growing diversity of the nation's students, including students with disabilities in general education classrooms, school districts and state departments of education face an urgent need for assessment tools that allow them to effectively and accurately report the proficiency level for all of their students. Moreover, results of these tools should be available at a speed that will allow them to modify instruction quickly and continue to make the changes necessary to meet proficiency levels. In addition, tools for accommodations specific to the attributes of these learners are urgently needed.

This article solely addresses "accommodations" associated with assessments, as differentiated from assessment "modifications." Accommodations afford adjustments to the practice of the assessment and do not alter the construct being evaluated. When an assessment is "modified," the construct inherent to the assessment is changed. Given the purpose and design of two assessments, the same allowance (e.g., extending time limits) could be an accommodation to both or an accommodation for one (a test built to be a power test) and a modification (change the inherent nature of the trait on a test intended to be speeded) for the other.

Edward L. Meyen is a professor in the Department of Special Education at the University of Kansas and codirector of the e-Learning Design Lab. John Poggio is a professor in the Department of Psychology and Research in Education at the University of Kansas and codirector of the Center for Educational Testing and Evaluation. Soonhwa Seok is a doctoral student in the Department of Special Education at the University of Kansas. Sean Smith is an associate professor in the Department of Special Education at the University of Kansas.

The IDEA legislation of 1997 created a requirement that all students with disabilities must be provided access to the general curriculum and accounted for in state and district assessment and accountability systems (Elliott, 2003; U.S. Department of Education, 2004). In addition, IDEA of 1997 states that students with disabilities must be allowed accommodations necessary to enhance instructional access as well as participation in statewide assessments (Elliott; U.S. Department of Education). To meet these requirements, testing must be customizable and individualized to assess, monitor, and instruct all students. Further, accommodations must be easily applied to all state- and district-level assessments.

For students with significant and developmental disabilities, alternative assessments and alternative achievement standards are allowed in meeting NCLB requirements. But only 1% of students assessed in a district may be included in Adequate Yearly Progress (AYP) documentation. Although each state defines AYP uniquely—which is relative to the rate of improvement—in the final analysis, by the year 2014, 100% of the students are to score at “proficient” or above on each of the state’s assessments in reading and in mathematics (Batt, Kim, & Sunderman, 2005).

Despite the reasonableness of the limit imposed by the 1% restriction, this provision has merit by at least moving the field to address the needs of this population of students. For the diverse population of students with high-incidence

disabilities, statewide assessments remain largely undifferentiated based on their needs or capacity. Accommodations, while of value, focus on test-taking behaviors (e.g., use of calculators) and practices (e.g., more time) but do not address curriculum alignment or item design to measure standards-based performance. In addition, they do not provide for tailoring assessments to the knowledge and skill sets of individual learners (McLaughlin & Nolet, 2005). Thus, Computerized Adaptive Testing (CAT) is compatible with curriculum alignment. The CAT also allows students with disabilities to demonstrate what they know in an efficient manner within the context of standards for all learners.

LITERATURE REVIEW

According to Fuchs (2000), accommodations refer to changes in standardized test conditions introduced to “level the playing field” between students with and without disabilities. In brief, the purpose of providing accommodations to students with disabilities is to allow them to achieve scores that are valid—not necessarily optimal scores. Thus, valid accommodations produce scores for students with disabilities that measure the same constructs that standard assessments measure in individuals without disabilities.

On one hand, disallowing valid accommodations prevents students with disabilities from demonstrating their competence. On the other hand, overly permissive accommodation policies inflate scores and thereby reduce pressure on schools to work hard to increase learning for students with disabilities. (Fuchs, 2000, p. 4)

A wide variety of accommodations can be found. In their review of the literature, Thompson, Blount, and Thurlow (2002) categorized 11 types of accommodation into one of four types:

1. Presentation
2. Response
3. Setting
4. Timing/scheduling

Presentation accommodations include (a) administering the test orally, (b) changing the test content by simplifying the language, or (c) changing the test format by using Braille or large print. *Response accommodations* include (a) allowing students to write directly into the test booklet, (b) dictating their answers to a scribe, or (c) using an electronic device to record their responses. *Setting accommodations* typically involve administering tests to students individually or in a separate room. Finally, *timing/scheduling accommodations* allow extra, unlimited time or breaking up the test administration into separate sessions (Sireci, Li, & Scarpati, 2003).

In a review of the literature on the effects of test accommodation on test performance, Sireci et al. (2003) referred to the literature in this area as “vast and passionate,” with a great deal of division among researchers on how they view accommodations. Some argue against test accommodations in the name of fairness to the majority of examinees who

FOCUS ON EXCEPTIONAL CHILDREN

ISSN 0015-511X

FOCUS ON EXCEPTIONAL CHILDREN (USPS 203-360) is published monthly except June, July, and August as a service to teachers, special educators, curriculum specialists, administrators, and those concerned with the special education of exceptional children. This publication is annotated and indexed by the ERIC Clearinghouse on Handicapped and Gifted Children for publication in the monthly *Current Index to Journals in Education* (CIJE) and the quarterly index, *Exceptional Children Education Resources* (ECER). The full text of *Focus on Exceptional Children* is also available in the electronic versions of the *Education Index*. It is also available in microfilm from Xerox University Microfilms, Ann Arbor, MI. Subscription rates: individual, \$36 per year; institutions, \$48 per year. Copyright © 2006, Love Publishing Company. All rights reserved. Reproduction in whole or part without written permission is prohibited. Printed in the United States of America. Periodical postage is paid at Denver, Colorado. **POSTMASTER:** Send address changes to:

Love Publishing Company
Executive and Editorial Office
P.O. Box 22353
Denver, Colorado 80222
Telephone (303) 221-7333

Susan T. Warhover
Editor

Stanley F. Love
Publisher

must take the tests under stricter, standardized conditions. Others argue that test accommodations are the only way to validly measure the knowledge, skills, and abilities of significant numbers of students. This matter of choice is an empirical question that can and must be addressed, not left to passion, beliefs, and emotions.

PERFORMANCE EXPECTATIONS

In addition to policies governing assessment of academic performance as a measure of effective schools, the theory of test design, and the accommodation of diversity among learners, the standards-based curriculum reform has introduced another variable that influences the potential consequences of statewide assessments for students with disabilities. Thus, the establishment of standards-based curricula across subject fields is based on the principle that performance expectations are increased, and that increased expectations, in turn, will result in more intense instruction and higher student achievement (Gersten & Baker, 2001).

National or state curriculum standards offer the advantage of providing direction for what should be taught, and when, to vast populations of students governed by elected policy-making bodies. Further, assuming that these standards are implemented as specified in the most recent mandates, they offer test designers specifics and needed guidance into the nature of items that reflect the curriculum standards. They also provide indicators as to when the skills, concepts, and knowledge associated with the standards will be taught and, subsequently, when students should be able to demonstrate competence (Poggio, 2001). For example, test designers must align items with the curriculum standards for what students should know about science in the sixth grade. As a result, tests are designed to assess student achievement at specific grade levels under the assumption that students at these grade levels have been taught (opportunity to learn) what will be assessed.

For the majority of students and schools, this assumption is probably true. Thus, states and districts have made major investments in curriculum planning centered on standards-based curriculum-reform initiatives, and teachers have invested time, and sometimes money, in professional development to better understand the standards and their professional responsibilities to students (King-Sears, 2003).

The provisions included in the reauthorization of the IDEA of 1997 make it reasonable to assume that standards-based curriculum reform is addressing the needs of students with disabilities. For example, IDEA requires that the individualized education program (IEP) for each student receiving special education services must include (a) statements describing how the student's disability affects his or her involvement and progress in the general curriculum; (b) measurable goals to enable the student to be involved in and progress in the general curriculum; and (c) the services, program modifications, and supports necessary for the student to be involved in and progress in the general curriculum (20 U.S.C. § 1400 et seq., 1997; 2004).

Although the current practice is for students from high-incidence disability groups to participate in inclusive settings, emerging evidence suggests that these students are not benefiting from this environment and related instruction (Palmer, Wehmeyer, Gipson, & Agran, 2004). Thus, despite their physical presence in the general education classroom, they often receive a differentiated curriculum based on their learning needs, and this precludes their being taught or learning the skills, concepts, and knowledge directly associated with the curriculum standards for the grade level of their peers and upon which statewide assessments are based (Deshler et al., 2001). Nevertheless, they are expected to participate in state- and district-level assessments. When justified, teaching "enabling skills" (skills that have to be in place to achieve grade-level targeted skills) will not yield a valid score for the students taking grade-specific content standards assessment.

Not surprisingly, data from statewide assessments indicate that students from high-incidence disability groups perform significantly below their nondisabled peers. In fact, a sizable proportion of students with disabilities have failed statewide assessments in several states (Olson, 2000a). For example, in Indiana in 1997, 21% of the students in special education, including students with learning disabilities (LD), failed the exam to graduate from high school, which prompted a class action lawsuit against the state (Olson, 2000b). Similarly, in California, 82% of students with disabilities (including students with LD) failed the language arts section of the 2001 high school exit exam required to receive a high school diploma (Egelko, 2002). This also prompted a lawsuit against the state (Asimov, 2003).

And in New York, only 29% of students with disabilities scored at or above Level Three (i.e., at or above state standards) on the 2002 Elementary-Level English Language Arts (ELA) Exam, compared to 66% of students without disabilities. Worse yet, at the secondary level only 9% of students with disabilities scored at or above the state standards on the 2002 Middle-Level ELA Exam, compared to 50% of the students without disabilities (New York State Education Department, 2003). In addition, the National Center on Educational Outcomes (NCEO) reported that in 17 states, a substantially smaller percentage of special education students than general education students met state standards (Ysseldyke et al., 1998).

In short, continuing state reports indicate that there is a group of students who have not yet been exposed to grade-level curriculum, making it impossible to offer fair or valid assessment for these students. In response to this argument, 14 states were testing students below the grade level in which they were enrolled in school during 2000–01 (Thurlow, Minnema, Bielinski, & Guven, 2003).

THE PROBLEM

To be able to demonstrate what they know more accurately, students with disabilities placed in general education classrooms need assessments designed specifically to

sample a broader range of understanding of curriculum standards tailored to their performance level. If students cannot demonstrate what they know (because they have not been taught the assessment tests skills), the test results will represent an inaccurate assessment of what they have learned, and, further, the actions the district takes to provide appropriate instruction may be misdirected.

In addition, action taken at the district, state, and/or federal level based on the overall performance of students in a district may be misguided, resulting in serious consequences for all students. For example, students scoring below "proficient" may need curricular interventions that are not appropriate for all students.

It is not a question of whether students with disabilities should be required to participate in statewide assessments or whether districts should be required to include these students' performance data when reporting AYP, or if students with disabilities should be held to high expectations or the same curriculum standards. The question is: How can we ensure that statewide assessments allow all students, including students with disabilities, to demonstrate their knowledge and skills on the curriculum standards being taught (opportunity to learn) and then assessed? Clearly, current methods are inadequate and have serious negative consequences for districts, schools, and, ultimately, students with disabilities.

THE SOLUTION

We will present a technology-based option that allows construction of tests tailored to the knowledge and skill attributes of individual examinees. For instance, technology can make available a wide array of accommodations that are content free but designed to enhance examinees' ability to demonstrate what they know about the content over which they are being examined. Computerized testing offers additional advantages, including the ability to report results within a shorter timeframe—one of the specifications of NCLB (Kingsbury & Hauser, 2004).

Consider the two scenarios below. The first scenario illustrates current assessment challenges, and the second offers a potential solution.

Scenario 1

A medium-sized rural school district in the Midwest has just concluded its mandated annual state assessment of seventh-grade proficiency in science. The state selects and mandates the tests. The students spent four class periods taking the tests, preceded by four weeks preparing by practicing test-taking skills and working on sample problems and items. At a later time, students will complete assessments across other subjects such as math, reading, and social studies.

All these assessments are designed to provide teachers, administrators and parents/guardians a snapshot of how well students are learning, how well teachers are teaching, and whether the district is making AYP. Following three months

of data capture, cleanup and psychometric analyses, the state department of education has scored the assessments and returned the results to the district.

Once again, the results are not what district stakeholders had hoped. For the third straight year, district scores failed to improve, and for the third straight year, revised instructional programs failed to deliver the promised improvements. Specifically, students with disabilities performed two to four years below their nondisabled peers despite special efforts to engage them in science instruction and improve their test-taking behaviors. The teachers are especially disappointed because their program for students with disabilities is known for its quality; indeed, many families have moved to the district because of its high reputation for serving students with disabilities.

The teachers argue that one of the reasons students with disabilities do so poorly in the content areas is their limited reading skills. Further, they believe that the test items do not allow these students to demonstrate what they know, and that students consequently lose interest in the tests quickly. Of equal importance, teachers and administrators are dissatisfied with the lack of meaningfulness of the results. That is, now that the assessment results are available, students have long since moved on to grade eight, and instructors are already engaged in teaching the next group of entering seventh-graders, so the results do little to guide or support instructional planning.

Local teachers and administrators do not believe that students with disabilities are benefiting from the assessments, yet they want assessment data on the students. In the end, the school and the district are made vulnerable by the poor performance of students with disabilities on state assessments. The end-of-the-year assessment ritual once again was the bearer of bad news, but the results arrived too late to serve an instructional purpose.

Scenario 2

A neighboring Midwest urban school district has concluded its annual assessment of seventh-grade science. Unlike their rural counterparts, however, teachers, administrators and parents/guardians already have a sense of how the students will score on the state assessment. That is because the district has adopted a new Computer Adaptive Test (CAT) measure of science achievement, which provides for multiple measures of student growth throughout the school year. Given this ongoing feedback relative to individual student achievements, district teachers have been able to assess students' progress in science and been able to adjust instruction regularly to meet individual student needs.

CAT has been especially helpful for teachers who work with students with disabilities, as the program's features offer several advantages to these students. Specifically, CAT is designed as a "smart" test, so once a student fails an item, the test automatically selects an item, covering the same content but at a lesser level of difficulty. In essence, the CAT tests the

same content for all students but tailors the assessment to each examinee's knowledge level. Thus, the CAT also increases the presentation of more difficult items if the examinee demonstrates an ability that warrants higher-level items.

These customized features result in students with disabilities showing more interest in and motivation to take the tests and being able to complete the assessment in a much shorter time than tests traditionally offered via the computer. In addition, the state has structured the system so the results are available within 24 hours or less via the Internet. This allows teachers and parents to make sound instructional decisions on an ongoing basis about the needs of individual students instead of having to wait until the end of the year only to discover that students were not making progress.

The teachers have a much richer understanding of each student's understanding of science and can make more informed decisions about what instructional changes have to be made. As a result of using CAT, the performance of the students with disabilities continues to improve, and the district is meeting its AYP goals successfully. Success in this scenario is the result of the immediate return of specific results for individual students and groups, and tests that efficiently align the appraisal with the students' learning.

PROPOSED TECHNOLOGY-BASED SOLUTION

Current technologies provide a basis for making Scenario 2 a reality for any school or district. First, though, we must address the primary question of how best to design statewide tests and associated analyses and report processes that maximize the ability of examinees who have high-incidence disabilities. The CAT makes it feasible to determine more precisely what accommodations represent added value for students with high-incidence disabilities.

Application of the CAT model may be viewed as an accommodation in itself because of its ability to tailor assessment effectively to the needs of individual learners, as well as its motivational effects on students' test-taking performance when items are aligned with their knowledge and skill sets. Thus, once the CAT is employed in statewide testing, all learners will benefit, including those with disabilities, English Language Learners, and advanced learners.

CAT models previously have not been applied at a significant scale or focused on students with disabilities, even though much research has extolled the benefits of computerized testing. These benefits include more precise measurement; ease of administration, scoring, and access to results; the ability to test at any time; and the potential for adaptive assessment tailored to individual examinees, thereby maximizing their validity on a case-by-case basis (Dunkel, 1999; Green, 1988; Green, Bock, Humphrey, Linn, & Reckase, 1984; Tonidandel, Quinones, & Adams, 2002; Thorndike, 1997; Wainer, 2000).

Even though the move toward computerized models in large-scale testing is logical and practical, several measurement issues must be considered, including student behavior

and performance. Although research on paper-and-pencil testing (P&P) has provided guidelines and models for measurement, the literature in this area does not fully apply to the computerized testing setting because of the inherent differences between the two modes (Glasnapp, Poggio, Poggio, & Yang, 2004). This applies even more so to CAT.

CAT FRAMEWORK: AN OVERVIEW

The following provides a more detailed description of the rationale behind and the technical underpinnings of CAT. Briefly, the traditional, fixed, P&P assessment presents the same number of questions to each test-taker without considering how well the person is performing on the assessment. Scores from this type of test usually depend on the number of questions answered correctly. That is, the more a student knows, the more questions he or she is likely to answer correctly.

A form of computerized assessment referred to as "computerized fixed-form" or "computerized fixed-length" tests is available. Except for the fact that the questions are presented on a computer screen, these devices are equivalent, relative to coverage, to P&P assessments. The same time limits apply, students can revise their answers, omit or pass over questions and, when desired, return to questions they have answered already, and so on. For this article, we view all such devices as traditional tests. The one distinguishing feature of the computerized, fixed-form assessment compared to its sibling P&P assessment is the notable ability to provide immediate results and, potentially, more accurate results.

In essence, a CAT works like a good oral examination. The algorithm first presents a question of moderate difficulty. In more sophisticated CAT systems, the initial question is determined based on a teacher's prior expectation of the student's likely performance level or a prior measure of achievement. When the student answers, the question is scored immediately. If the answer is correct, the test statistically estimates the examinee's knowledge level as higher than previously estimated and finds and presents a question that matches that higher skill level. If, however, the first question is answered incorrectly, the opposite occurs.

After presenting the second question, the test waits for the answer. Once the test-taker answers, the computer scores it. If correct, the statistical computing algorithm reevaluates the student's knowledge as even higher; if incorrect, the software reestimates the knowledge level as lower and searches for a third question to match the new estimate of overall knowledge. The process continues like this, with the test methodically and gradually locating the examinee's "true" knowledge level for the content being tested.

Thus, the score that serves as an estimate of achievement or standing becomes more accurate with each question answered. The test ends (a) when the accuracy of the estimate reaches a statistically acceptable level, (b) when a maximum number of test items have been presented, and

(c) the content coverage of the questions presented is deemed representative of the tested domain.

Figure 1 shows the estimation of a test-taker's achievement after answering each of 10 test questions. Notice how the knowledge level is estimated lower when questions are answered incorrectly (Questions 3, 6, 8, and 10). The dotted vertical lines indicate the amount of error associated with the achievement estimates (and correspondingly, the degree of confidence in the score). As more questions are presented and answered, this error in estimating the student's knowledge level decreases.

The CAT is expected to end when the amount of measurement error around the knowledge-level estimate reaches an acceptable level. Low levels of measurement error indicate that the test likely would produce a similar score if it were readministered. A CAT's estimate of measurement error at a score or decision point (e.g., Basic/Proficient) in the central area (e.g., P15 to P85) of a score distribution is very comparable to measurement error on fixed-form tests. Because it is not known ahead of time exactly when the test will end, a CAT usually presents a variable number of questions to examinees of different skill levels, although minimum and maximum numbers of questions are typically set.

Of particular relevance to the current discussion is that in test scores derived from a CAT, a person with less prior evidence of learning mastery is able to answer the same number of questions correctly as a person who typically scores higher. Comparing the questions answered correctly for both persons would reveal that the latter was able to answer more difficult questions correctly, and therefore should receive a higher score. That is exactly what happens. The score is not based on the number of correct answers but, rather, on the level of difficulty of the questions answered correctly.

The score is computed based on the principles of item response theory (IRT) that was proposed initially by Lord

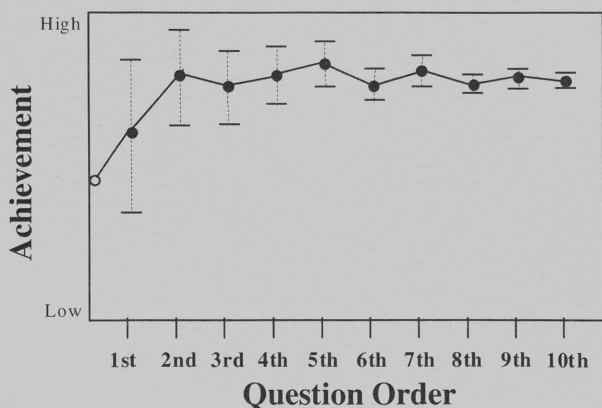


FIGURE 1
A Typical Pattern for a CAT

(1980). The formula that calculates the final score converts it to a scale that becomes interpretable. Just as with a traditional P&P, a classification likelihood score also may be determined—such as Exemplary, Advanced, Proficient, Basic, and Unsatisfactory. In fact, when a CAT is prepared to attain classification versus score accuracy, the efficiency and accuracy benefits of a CAT improve dramatically (Young, Shermis, Bruten, & Perkins, 1996).

As mentioned, too often students with disabilities in general education curriculum are not performing at grade level or even experiencing the same curriculum that is offered to their nondisabled peers in the same classroom. Yet, when these students participate in assessments required by NCLB, they are subjected to tests designed to assess what presumably is taught in the general curriculum at grade level. This potentially places them at a great disadvantage. The CAT has the advantage of being able to assess a wide range of knowledge and skills and to tailor the assessment items to the individual examinee's ability.

Moreover, for any given person, some questions on a traditional P&P test are far too easy and some are far too difficult. The answers of highly skilled test-takers to easy questions do not meaningfully inform instruction for them as most people answer the easy questions correctly. Similarly, for test-takers who have not mastered the material, failing to answer the difficult questions correctly also tells us very little. Instead, if a traditional test were able to discover the level, on a scale from easy to difficult, at which the examinee begins to find the questions to be beyond his or her readiness level, a score could be derived for that level and the test would be more effective. The proposed CAT does just that!

In addition, a CAT tailors itself to the knowledge and skill level of the individual test-taker (Sands, Walter, & McBride, 1997). By taking into account how each person taking the same CAT answered previous questions, the test adjusts to give the lower-achieving student a set of questions different from those given to another student with a different achievement level. Thus, the former will see relatively easy questions, whereas the latter will see more difficult questions. Both students may answer the same percentage of questions correctly, but because the higher performing student can answer more difficult questions correctly, he or she will get a higher score.

Figure 2 offers a general, simple framework for understanding how a CAT works. Each box represents a different test question. As illustrated, giving a correct answer to a question results in a more challenging question being presented, whereas an incorrect response results in a somewhat easier question being presented next. After exposing students to a more restricted set of questions, the assessment terminates and a score is assigned.

Although computerized testing is relatively new to education—in particular, large-scale assessment program offerings—it is not a new enterprise (Parshall, Spray, Kalohn, & Davey, 2002). Software has been available for some time,

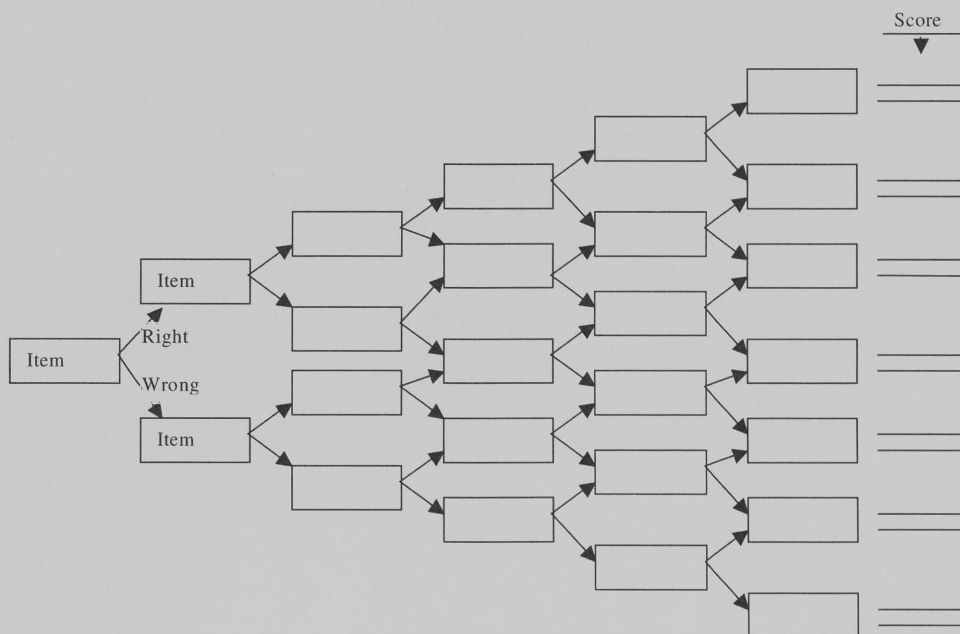


FIGURE 2
Computer Adaptive Testing Framework

and fixed-form and adaptive testing have been dominant methods of assessment in business and industry for a decade or longer. Companies such as Microsoft, Cisco Systems, and Novell, as well as several allied health professions, have been carrying out widespread computer-based testing successfully for years (Wainer, 2000).

In education within the past decade, Educational Testing Service (ETS) has pioneered both fixed-form and adaptive computer-based testing in many of its most recognized programs (GRE, GMAT, SAT, Praxis, etc.). Further, within the past couple of years, seven states have delivered their large-scale assessments online as computerized assessments—Idaho, Oregon, South Dakota, North Carolina, Virginia, Indiana, and Kansas—and

... an increasing number of states including Arkansas, Georgia, Idaho, Kansas, Kentucky, Maryland, North Carolina, Oregon, South Dakota, Texas, Utah, and Virginia are pursuing comparability studies or have already implemented computer tests as part of their K–12 assessment programs. (Paek, 2005, p. 15; Bennett, 2003; Olson, 2003)

More states (e.g., New Jersey, Delaware, Tennessee, and Utah) are planning and acting on the move to computerized online testing. What we can deduce from these K–12 education trends is that schools are acquiring the staff competencies to use online assessment effectively, staffs and students are accepting of computerized assessments, and many school environments have the capacity to engage in and carry out computer-based testing. Nevertheless, only two

state assessment programs—Oregon and Kansas—are systematically employing the CAT model (Center for Educational Testing and Evaluation in the University of Kansas, 2005; Tindal & Haladyna, 2002).

CONCLUSION

To accept accommodations in testing as the solution to achieving equity in statewide assessments for students with high-incidence disabilities is insufficient, if not unacceptable. The tests themselves must be designed to maximize the opportunity for all students to demonstrate what they know relative to the standards being assessed. For this to occur, tests must have the capacity to tailor items to the knowledge level of all examinees and, subsequently, efficiently assess what they know. This is particularly important given the mandates of NCLD, as merely assessing what students do not know does not inform instruction, nor does it motivate learners toward higher performance.

Test-construction theory and supportive technologies are in place to broadly implement CAT in large-scale assessments. Thus, advocates of equity for students with high-incidence disabilities in statewide testing should work with policy makers in their states to ensure that developers of statewide tests fully explore the applications of CAT. Likewise, the effectiveness of CAT in assessing the performance of students with high-incidence disabilities has to be researched.

FOCUS ON EXCEPTIONAL CHILDREN

The requirements of NCLB and IDEA should be leveraged to provide the conditions necessary to bring about the changes needed in statewide testing to ensure that all students are able to demonstrate what they know. Above all, the results of such testing should serve to inform instruction as teachers seek to make evidenced-based decisions while working to maximize student achievement.

REFERENCES

- Allbritten, D., Mainzer, R., & Ziegler, D. (2004, Winter). Will students with disabilities be scapegoats for school failures? *Educational Horizons*, 82(2), 118–168.
- Asimov, N. (2003, May 2). Exit exam too hard—reprieve likely. State may delay denial of diplomas to students who fail. *The San Francisco Chronicle*, p. A1.
- Batt, L., Kim, J., & Sunderman, G. (2005). Policy brief: Limited English proficient students: Increased accountability under NCLB. Retrieved February 7, 2006, from http://www.civilrightsproject.harvard.edu/research/esea/LEP_Policy_Brief.pdf
- Bennett, R. E. (2003). *Online assessment and the comparability of score meaning*. Princeton, NJ: Educational Testing Service.
- Center for Educational Testing and Evaluation in the University of Kansas. (2005). Kansas computerized assessment. Retrieved February 8, 2005, from <http://kca.cete.us/kca.html>
- Deshler, D. D., Schumaker, J. B., Lenz, B. K., Bulgren, J. A., Hock, M. F., Knight, J., et al. (2001). Ensuring content-area learning by secondary students with learning disabilities. *Learning Disabilities Research and Practice*, 16(2), 96–108.
- Dunkel, P. A. (1999). Considerations in developing and using computer adaptive tests to assess second language proficiency. *Language Learning & Technology*, 2(2), 77–93.
- Egelko, B. (2002, February 22). State must provide for disabled on test. *San Francisco Chronicle*, p. A21.
- Elliott, J. L. (2003). IDEA 2003: Reauthorization or Retrofit? *School Administrator*, 60(3), 28–30.
- Fuchs, L. S. (2000). Research report on reading. The validity of test accommodations for students with learning disabilities: Differential item performance on reading and mathematics tests as a function of test accommodations and disability status. Delaware Department of Education: #2. Retrieved June 16, 2004, from <http://www.doe.state.de.us/aab/Atch%205%20The%20Validity%20of%20Test%20Accommodations%20for%20Students%20with%20Learning%20Disabilities.pdf>
- Gersten, R., & Baker, S. (2001). Teaching expressive writing to students with learning disabilities: A meta-analysis. *The Elementary School Journal*, 101, 251–272.
- Glasnapp, D., Poggio, J., Poggio, A., & Yang, X. (2004, April). *Student attitudes and perceptions regarding computerized testing as a method for formal assessment*. Paper presented at annual meeting of National Council on Measurement in Education, San Diego, CA.
- Green, B. F. (1988). Critical problems in computer-based psychological measurement. *Measurement in Education*, 1, 223–231.
- Green, B. F., Bock, R. D., Humphrey, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21, 347–360.
- Kingsbury, G., & Hauser, C. (2004). Computerized adaptive testing and No Child Left Behind. Retrieved February 8, 2006, from <http://www.ecs.org/html/IssueSection.asp?issueid=195&subissueid=101&ssID=0&s=Selected+Research+%26+Readings>
- King-Sears, P. (2003). Traveling the road of research-to-practice: Scaling up for the long journey. Paper presented at 25th International Conference on Learning Disabilities Conference, Seattle.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- McLaughlin, M., & Nolet, V. (2005). *Assessing the general curriculum* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- New York State Education Department. (2003). Understanding your school report card 2003. Retrieved February 5, 2006, from <http://www.emsc.nysed.gov/reprcd2003/information/elementary/guide.html>
- Olson, L. (2000a). Worries of a standards “backlash” grow. *Education Week*, 30, 1–13.
- Olson, L. (2000b). Indiana case focused on special education: Suit challenges high stakes testing. *Education Week*, 38, 1–15.
- Olson, L. (2003). Legal twists, digital turns: Computerized testing feels the impact of “No Child Left Behind.” *Education Week* 12(35), 11–14, 16.
- Paek, P. (2005). Recent trends in comparability studies. Retrieved February 7, 2006, from http://www.pearsonedmeasurement.com/downloads/research/RR_05_05.pdf
- Palmer, S. B., Wehmeyer, M. L., Gipson, K., & Agran, M. (2004). Promoting access to the general curriculum by teaching self-determination skills. *Exceptional Children*, 70(4), 427–439.
- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.
- Poggio, J. (2001). The renewal of acre: Planning a revision design to support religious education teaching and learning. *Momentum, February/March*. Retrieved February 8, 2006, from <http://www.ncea.org/programmevents/rea/acre-renewal.pdf>
- Sands, W., Walter, B., & McBride, J. (1997). *Computerized adaptive testing*. Washington, DC: American Psychological Association.
- Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature (Center for Educational Assessment Research Report No. 485)*. Amherst: University of Massachusetts, School of Education; Retrieved June 16, 2004, from <http://education.umn.edu/NCEO/OnlinePubs/TestAccommLitReview.pdf>
- Thompson, S., Blount, A., & Thurlow, M. (2002). *A summary of research on the effects of test accommodations: 1999 through 2001* (Technical Report 34). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved June 15, 2004, from <http://education.umn.edu/NCEO/OnlinePubs/Technical34.htm>
- Thorndike, M. R. (1997). *Measurement and evaluation in psychology and education*. Columbus, OH: Merrill.
- Thurlow, M., Minnema, J., Bielski, J., & Guven, K. (2003, June 28). Testing students with disabilities out of level: State prevalence and performance results (Out-of-Level Testing Project Report 9). Retrieved June 29, 2004, from <http://education.umn.edu/NCEO/OnlinePubs/OOLT9.html>
- Tindal, G., & Haladyna, T. M. (Eds.) (2002). *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation*. Hillsdale, NJ: Erlbaum.
- Tonidandel, S., Quinones, M. A., & Adams, A. A. (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test-takers’ reaction. *Journal of Applied Psychology*, 87(2), 320–332.
- 20 U.S.C. § 1400 *et seq.* (1997). Individuals with disabilities education act.
- 20 U.S.C. § 1400 *et seq.* (2004). Individuals with disabilities education improvement act.
- U.S. Department of Education. (2002). No Child Left Behind Act (P.L. 107–110). Retrieved February 5, 2006, from <http://www.ed.gov/nclb>
- U.S. Department of Education. (2002). To assure the free appropriate public education of all children with disabilities: Individuals with disabilities education act, section 618. Retrieved February 2, 2005, from http://www.pluk.org/Pubs/Fed/IDEAreport_2002_4.8M.pdf
- U.S. Department of Education. (2004). IDEA ’97 Amendments: Final regulations. Retrieved February 5, 2006, from <http://www.ed.gov/policy/speed/reg/regulations.html>
- Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ysseldyke, J., Thurlow, M., Langenfield, K., Nelson, J. R., Teelucksingh, E., & Seyfarth, A. (1998). *Educational results for students with disabilities: What do the data tell us?* (Technical Report 23). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Young, Y., Shermis, M. D., Bruten, S. R., & Perkins, K. (1996). From conventional to computer-adaptive testing of ESL reading comprehension. *System*, 24, 23–40.