The Impact of High-Stakes Testing for Individuals with Disabilities: A Review Synthesis

Richard Boon, Ph.D.
The University of Georgia

Debbie Voltz, Ph.D.
The University of Alabama-Birmingham

Carl Lawson, Sr.
Chicago State University

Michael Baskette, M.Ed.
The University of Georgia

Abstract

The purpose of this paper is to provide a review of the literature evaluating the impact of high-stakes testing for students with learning disabilities. In this review, 30 studies were obtained, reviewed, and synthesized. Findings discuss the definition and description of high-stakes testing, origins of high-stakes testing for students with disabilities, educational outcomes, effects on curriculum, and impact of high-stakes testing for students with learning disabilities. Results and limitations are discussed in relation to instructional practice and future research issues to extend the current literature findings.

America has a long tradition of weaving assessment into school improvement equations. Predictably, for several years accountability for test scores has been viewed as key to productive educational improvement (Stiggins, 1999). Landau, Vohs, & Romano, (1998) and McGrew, Spiegel, Thurlow, Shriner, & Ysseldyke, (1994) found school accountability reform has raised major issues concerning the educational treatment of students with disabilities and their academic achievement. The National Council on Disability (1993) in its report to the President and Congress noted that for the years 1986 through 1989 the proportion of students with disabilities who dropped out or left school for undetermined reason increased from 25% to 27% and 12% to 18%, respectively.

McGrew, Thurlow, and Spiegel (1993) noted that across the country, 40% to 50% of students with disabilities of all school age are excluded from various large-scale assessments. More recently, Thurlow (2001) reported that 31%, 20.7%, and 15.1% of students with disabilities in 4th, 8th, and 10th grade, respectively, were not tested in Nevada. In other states, inclusion in the assessment system does not always mean that students' scores are included in the average used to determine rewards or sanctions. For instance, in 2001 Louisiana reported that the scores of

94.3% of third graders, 94.2% of fifth graders, 93.9% of sixth graders, 92.3% of seventh graders, and 88.8% of ninth graders were excluded from the school averages. Encouraging such practices implies that the learning achievement and progress of students with disabilities do not count.

Landau et al (1998) and Ysseldyke, Thurlow, Kozleski, & Reschly (1998) have documented the need for a more inclusive outcome assessment system. McGrew et al (1994) reported the results of a survey of state assessment data that, while most states affirmed that they include some students with disabilities in their accountability system, only six out of 50 states were able to provide data about their large-scale assessment. McGrew et al (1993) suggests that many students with disabilities have been excluded from large-scale achievement data in many states across the nation. Considering that school reform activities use measurable indicators from large-scale high stakes assessment as index of progress, McGrew et al (1994) concluded that it is imperative that states implement uniform strategies for increasing and documenting the inclusion of these students in state data collection programs. They have advocated the need for a more holistic school reform and standard-based accountability system that promotes systematic efforts to include all students with disabilities in school outcome measures. The aforementioned authors also suggested that the use of high-stakes assessment in educational decisions would lead to better outcomes for all students, including students with disabilities.

In the late 1980s testing was promoted as a way of ensuring that educational standards were met and state and district-wide large-scale assessment was viewed as a way to hold schools accountable for all students' learning outcomes. Popham (1987) postulated that only if the stakes are high, meaning if there is something valuable to gain or lose, will teachers and students take education and tests seriously and work hard to do their best. Landau, et al. (1998) have noted that including students with disabilities in assessment sends the message that schools are accountable for all students teaching higher levels of learning. Conversely, Allington & McGill-Franzen (1992) reported that, in some instances, high-stakes testing rewards harmful instructional practices rather than school improvement. Langenfeld, Thurlow, and Scott (1997) examined the effects of high stakes testing for students and concluded that administering tests that have important consequences for students, teachers, and the school could adversely impact instruction. Despite the apparent interest, very few investigations have been conducted in the area of high stakes-testing on students with disabilities in general. A review of the literature in the area of high-stakes assessment revealed very few research studies that examine high stakes assessment in relation to their potential impacts on students with mild disabilities on one hand, and students with severe and profound disabilities on the other hand.

In spite of the noticeable lack of research supporting the effectiveness of inclusive assessment, an increasing number of states across the nation are implementing high stakes assessment with students with disabilities (Heubert, 2000; Thurlow, 2001). As more inclusive large-scale assessment is becoming the standard practice, there is an urgent need for more research that focuses on the specific impact of high stakes tests on students with mild disabilities (Langenfeld, Thurlow, & Scott, 1997). Investigating the effects of high stakes testing on students with learning disabilities has never been so important, especially as educators and legislators are trying to better interpret and use assessment results.

High-Stakes Testing: Definitions and Descriptions

High sakes tests are also called "exit exams, certification exams, or competency exams" (O'Neill, 2000). High stakes testing means that a test is given to students and the test score is the sole measure used to make crucial decisions about students, teachers, and schools regardless of previous and future performance (Tingley, 1999). Students' scores on a test can "bring public praise or financial rewards" or "public embarrassment and heavy sanctions" to the school during high-stakes testing (AERA, 1999). Individual students could be placed in honors classes or programs for the gifted. Similarly, if students score low or do not meet the standards they might not be promoted to the next grade or will not graduate from high school. In some instances, three-fourths of the school staff could be replaced (Langenfeld, et al., 1997). Orfield and Wald (2000) reported that high-stakes testing policy might link the score on one test to teachers' and principals' salaries and tenure decisions. Obtaining a low score on the test also increases the likelihood that students might be rejected by a particular college (Ransom et al., 1999) or for a particular employment opportunity. Elsewhere, high-stakes tests imply that an individual student's score is used to determine student's needs, and whether he or she will be allowed to enroll in a certain academic program. However, many individuals and institutional viewpoints consider such a practice unacceptable. Ducharme and Ducharme (1998, p.83) noted that the "current trend and emphasis being promoted across the nation and several states is potentially dangerous and tragic." The American Educational Research Association declares that "decisions that affect individual students' life chances or educational opportunities should not be made on the basis of test scores alone" (AERA, 2000).

Origin of High-Stakes Testing for the Students with Disabilities

Before the passage of public law 94-142 (Education of All Handicapped Children Act [EHCA]) in 1975, the education of students with disabilities was not mandatory in the United States. The public school system was neither required to accept them in the classroom nor to provide an "appropriate" education that maximizes their potential. As a result of such discriminatory policies, many students with disabilities were institutionalized and others were simply secluded indoors (Ysseldyke et al. 1998). Many advocacy groups struggled to provide students with disabilities equal access to public school buildings and appropriate education in the 1970s. Furthermore, with the passage of P.L. 94-142 in 1975, the education of all students with disabilities became mandatory, free and appropriate (Yell, 1997). Unexpectedly, twenty years after the 1975 landmark act, special education programs in general have been far from meeting their intended expectations (Danforth & Rhodes, 1997). Not long ago, research on school reform noted the pervasive and systematic exclusion of students with disabilities from the national data analysis used to report educational improvement. McGrew, et al (1993) state that the systematic exclusion of students with disabilities in data analysis characterizing the period between the 70s and early 90s adversely affected the educational outcomes and the general attitude toward the employability and placement of such students.

Current Research on High Stakes Tests and Students with Learning Disabilities

Until recently, studies on the effects of high stakes assessment on students with disabilities were practically not available in the literature. The scarcity in the research base could be due to the fact that very few states in the nation included students with disabilities in their assessment data collection (McGrew et al., 1994). It was not until 1997 that the amendments of IDEA required that students with disabilities be included in accountability programs. In high stakes assessments, all students with learning disabilities are not subjected to the same rules and regulations.

Students with learning disabilities represent the sub-group of students with special needs ages 6-21 that perform below their cognitive abilities in one or more academic areas. These students are referred to as having mild disabilities because most of their needs and characteristics go undetected until they reach school age (Henley, Ramsey, & Algozzine, 1996; Thurlow, Elliot, & Ysseldyke, 1998). The prevalence of students with learning disabilities is difficult to estimate due to the different eligibility criteria requirements used in different states. Henley et al (1996) noted that in most cases these students are enrolled full time within regular classrooms with accommodations or receive special services in the resource room one or more periods a day. Consequently, schools use different methods of assessment to obtain a comprehensive picture of their achievement. These methods include traditional assessment with or without accommodation in most of the cases, supplemented by alternate assessment in very few cases.

The use of statewide and nationwide standardized test scores to measure educational outcomes for students with disabilities have increased over the last two decades. This increase is a result of major legislative reforms including Goals 2000, School-to-Work, Improving America's School Act, and the Individual with Disabilities Education Act (IDEA) (National Center on Education Outcomes [NCEO], 1996). As early as 1980, high stakes tests for high school exit were mandated for students with disabilities in Maryland, Kentucky, North Carolina, and Texas. For instance, the Maryland School Performance Program (MSPP) accountability system was established in 1989 by the State Board of Education as a "vehicle to move toward a high quality educational system for all of Maryland's students in the 21st century (NCEO, 1996). The MSPP requires that students with disabilities be included in state and district accountability systems. For any student to be excluded from this large-scale testing he/she must be a second semester senior transferred from out-of-state, a first time Limited English Proficient student, or not pursuing the Maryland Learning Outcomes which included scores in reading, writing, language usage, mathematics, science, and social studies. Similarly, the Accountability Based Curriculum (ABC) system in North Carolina (Jones, 1999), the Texas Assessment of Academic Skills (TAAS) (Natrello & Pallas, 1999), the Accountability testing in Kentucky (Stecher & Barron, 1999) and other systems share a common feature: they are totally inclusive. The accountability systems of ABC, TAAS, and Kentucky include students with and without disabilities who are subjected to the same, or slightly different, academic standards.

The 1990s have witnessed a significant impulse in the history of inclusive assessment. President Clinton's 1997 State of the Union address to the nation proposed that all students take a national test of reading in fourth grade and mathematics in eighth grade. An even more important impetus for increased focus on inclusive assessment occurred on June 1997 when Public Law 94-142 was reauthorized. Public Law 105-17 (IDEA 1997) included the requirement that students with disabilities have access to the same high standards and general education curriculum as their

non-disabled peers (Yell, 1998). Public 1aw 105-17 also requires that students with disabilities be included in a large-scale assessment with accommodations and adaptations provided when and as needed.

In most states and districts, traditional assessment relies on criterion-referenced tests (Thurlow et al., 1998). The reason for this is that this type of test creates fewer challenges for accommodations and also allows teachers to measure students' performance against a specific criterion. It is a requirement of the law to provide students included in district and state accountability system, and eligible for traditional assessment, with the needed accommodation to "level the playing field" (Thurlow et al. 1998, p 29). Accommodations are changes in testing materials or procedures that enable students with disabilities to participate in an assessment in a way that allow abilities, rather than disabilities, to be assessed (Thurlow et al., 1998). Thurlow et al identified five main types of possible accommodations used in high stakes assessment settings. They include time accommodation, setting accommodation, scheduling accommodation, presentation accommodation, and response accommodation. Examples of accommodations are presented in **Table 1**.

The importance of the test accommodation is not always clear to everyone in cases of cognitive disabilities affecting learning. Controversy arises when it is believed that accommodations used with norm-reference test may change the nature of the test and in some cases significantly affects the meaning and interpretation of the students' scores (Phillips, 1994). It is still debated which accommodations preserve the meaningfulness of students' score. Advocates of test accommodation argue that providing students with disabilities with needed accommodations is fair. For them testing conditions should be altered for students with disabilities to compensate for neurological problems (Phillips, 1994). That is why some states using norm reference standardized tests (e.g. Kentucky, and Louisiana) provide students with various accommodations (Thurlow, 2001). Opponents of test accommodations often believe that some accommodations might be beneficial to students who receive them and invalidate the inference that can be made from students' performances. Not all students with mild disabilities are accommodated. Before accommodation is implemented for a student during testing it has to have been used previously during classroom instruction. When students with disabilities receive accommodations, information should be provided as far as when, what, and how it is done in the report of the test final.

In some instances, students with mild disabilities might be eligible for alternate assessment. Indeed, the 1997 Amendments of IDEA mandates that, no later than July 1, 2000, alternate assessment be an option for students who, due to the severity of their disabilities, cannot participate in the general large-scale assessment used by states and districts (IDEA Regulations, 34 C.R.F. s 300.138 (b) (1) (2) (3)).

Students who are eligible for alternate assessment might be tested on the basis of the state's content standards for all students. The content of the assessment and the strategies used to collect information on how well students are progressing toward the standards vary tremendously from one student to another and from one district to another. Thompson, Quenemoen, Thurlow, and Ysseldyke (2001) identified several forms of alternate assessment. They include performance-based assessment, authentic assessment, and "alternative" or portfolio assessment, the latter being defined as "a purposeful and systematic collection of students' performance assessment relative to standard" (p91). In either case teachers use observation, recollection, and record review to collect information on students' learning outcomes. When students take

alternate assessment or the regular test with accommodations, performance should be included in state report and "flagged" showing that a particular student, even though included in the accountability system, has taken a particular assessment (Kleinert, Kennedy, & Kearn, 1999; Thompson et al., 2001).

The purpose of this paper is to provide a thorough analysis of the research literature from 1990-2004 on the effects and impact of high-stakes testing on the special education curriculum, students, and teachers and the educational outcomes of students with disabilities and the school reform movement. Finally, through such a review researchers can gain greater insights into future efforts to, not only include more students with learning disabilities in participating in high-stakes testing, but to increase the probabilities of success on these tests.

Table #1

Table 1
Examples of modifications and interventions that can be used as accommodations on high-stakes testing.

Academically Related		
Study sheets	Waive time constraints	Peer assistance
Scheduled breaks	Lower/higher level material	Task/assignment sheet
Test read to student	Note taking aids	Reduce task (number)
Reduce task (length)	Small groups	Equipment
Alternative test	Multiple choice vs. essay	Highlight skills
Behaviorally Related		
Adult proximity	Preferential seating	Isolated area
Contract	Reinforcement	Management system

Literature Search Procedures

The following literature search procedures were employed to retrieve relevant articles. First, a computer-assisted search of four major databases was conducted including ERIC, PsycInfo, ArticleFirst, and Dissertation Abstracts. The descriptors used in the search were "testing, assessment, disabilities, mildly handicapped, and learning problems". Second, after an extensive evaluation of the relevant electronic and paper journal articles was completed the references of these articles were examined to determine if any other articles were available that had not come up in our initial search. Finally, a hand search of reference lists and table of contents of relevant journals was conducted. This search revealed 30 studies which met the criteria for inclusion in this review.

Criteria for Inclusion

The three main criteria for inclusion in this review include: (a) articles published from 1990 to 2004; (b) studies that examined the impact of high-stakes testing; and (c) individuals included in the study are classified as having a learning disability by the authors. Due to the numerous changes to major special education laws that have taken place since 1990 our search did not include articles published prior to this date. For the purposes of this review, studies were excluded when subjects were not classified as having a learning disability in the article.

Overall Study Characteristics

There were 30 studies that ranged in publication date from 1990-2004, and appeared in referred journals such as the Journal of Special Education, Journal of Learning Disabilities, Exceptional Children, Journal of Adolescent and Adult Literacy, The Reading Teacher, Education and Training of the Mentally Retarded, Remedial and Special Education, Journal of Teacher Education, Educational Researcher, School Psychology Review, Educational Evaluation and Policy Analysis, Applied Measurement in Education, Schools in the Middle, The Nation, Phi Delta Kappan, and Canadian Journal of Education.

Gronna, Jenskins, and Chin-Chance (1998) investigated the longitudinal performance of students with disabilities in a norm-referenced statewide standardized testing program during the years 1992 to 1996. The statewide study conducted in Hawaii, included 24,595 students without disabilities and 21, 411 students with disabilities in high-incidence categories (mild mental retardation [MMR], emotional impairment [EI], speech and language impairment [SLI], and specific learning disabilities [SLD]) who took the Stanford 8 without accommodation. Gronna, Jenskins, and Chin-Chance (1998) used a one-way analysis of variance with multiple-range post hoc Bonferroni tests to compare students with disabilities in Stanford 8 norm group with the population of students with disabilities in Hawaii. They reported that all students with disabilities scored lower means than the national normative group. The mean scores for students with MIMR, SLD, and EI is significantly different from that of the non-disabled students in all grades tested in reading and mathematics, from that of the non-disabled students in all grades tested in reading and mathematics, from that of the non-disabled students in all grades tested in reading and mathematics, whereas students with SLI were similar in performance to non-disabled population.

Hollenbeck, Tindal, and Almond (1998) conducted a pilot study for Oregon Department of Education to determine teachers' knowledge about allowed accommodations on statewide assessment and whether those accommodations were uniformly implemented across the state. One hundred and sixty six teachers randomly selected responded to the survey. Teachers' knowledge of accommodation was organized in four groups (strong, average, moderate, and weak high stakes decision power) and analyzed. The authors found that most teachers fell into the weak knowledge group (96.4%) and none (0%) in the strong knowledge group. The authors concluded that teacher's knowledge of accommodation was "limited enough to jeopardize the validity of score interpretation across the states for various subgroups for lack of test administration reliability" (p181). In addition, they reported that general education teachers reported use accommodations more often than special education teachers. The study also reveals that very few of the accommodations used in high stakes testing reflect universal agreement among respondents.

Kleinert, Kennedy, and Kern (1999) conducted a statewide survey of teachers involved in Kentucky's first alternate assessment and accountability system for students with moderate to severe disabilities. Two main research questions were investigated:

1) To what extent do teachers perceive benefits of including students in state and district alternate assessment accountability measures?

2) What are teachers' perceptions of the instructional impact of assessment on students' outcomes? Three hundred and thirty one teachers were surveyed.

The authors found that overall teachers mentioned frustration at the increased work involved in alternate assessment process. However, they believe that it is important to include all students in the state and district accountability system. When asked whether they perceive their students benefiting from being included, 52.9% of the teachers agree or strongly agree. In contrast, teachers were less positive about the impact of alternate portfolio on helping them leverage access to general education classes.

Fuchs et al. (2000) studied the effects of test accommodations. One hundred and eighty one fourth and fifth graders with LD and 184 fourth graders without LD participated in the study, which examined whether students with LD benefit from accommodations more than students without learning disabilities. The students were to complete four brief reading assignments under four conditions: standard, extended time, large print, student reading aloud. After analysis of student's outcomes the authors found that, for extended time and large print, students with LD did not benefit more than their counterparts without disabilities. Effect size for these accommodations was almost similar with the highest effect size for students with LD (.36 and .38 for extended time; .03 and .08 for large print). In contrast to this result, they found a statistically significant interaction for students reading aloud, showing that this particular accommodation may increase scores of students with LD and depress scores for students without disabilities.

Fuchs, Fuchs, Eaton, Hamlett, and Karns (2000) compared the effects of accommodations in three mathematics domains of curriculum based measurement (CBM); computation, concepts and applications, and problem solving for students with and without LD. The primary purpose of the study was to determine if students with LD benefit from specific accommodation. Participants in the study include 181 fourth graders without LD and 192 fourth and fifth graders with LD, all tested at third grade level because the study began in the Fall, and it was believed that in the Fall students' knowledge of fourth grade curriculum is still not mastered. Furthermore, each teacher of students participating in the study was instructed to determine for each student whether accommodation should be provided and which accommodation should be implemented. After running a mixed model of two-way analysis of variance (between-subject and within subject ANOVA), the authors found a significant difference indicating that students with LD do benefit from extended time in mathematics areas that require reading extended text and producing extended verbal, written answers. Fuchs et al used McNemar post-hoc test for dependent sample to measure teachers' decision-making skills regarding awarding accommodations and found that teachers over-awarded accommodations to students.

Kampfer, Horvath, Kleinert, and Kearns (2001) examined the amount of time and effort required on states' alternate assessment. The authors surveyed 206 special education teachers who had a student participating in the Alternate Portfolio assessment in Kentucky during the 1998-1999 school years. They reported that 66% of teachers stated that they spend and enormous amount of time preparing this type of assessment, on average between 25 and 35 hours outside of instruction per portfolio. Further, the authors asserted that teachers perceived "some" benefit for the students in participating in the Alternate Portfolio assessment.

Effects of High Stakes Testing on Curriculum, Teachers, and Students

Very few empirical studies have been conducted in the area on high stakes testing effects on students with disabilities. Studies conducted in regard to this issue, prior to the 90s, were predominantly position papers. In the frame of this paper, and due to the rarity of statistics in the area, both research articles and position papers are examined in the following review.

Lacina-Gifford and Kher-Durlabhji (1993) have identified three basic problems resulting from the use of high stakes tests with students. 1) They emphasized that the methods teachers use to insure good performance by students in high stakes tests do not secure learning. 2) At all grades the curriculum is narrowed and reduced to the content of the test. 3) The authors noted that the use of a single measure to determine students' future raises some ethical questions.

Wideen, O'Shea, Pye, and Ivany (1997) conducted a two-year case study to explore the relationship between high stakes testing and the teaching of science in two school districts randomly selected from the ten districts in British Columbia. The authors interviewed a total of 80 teachers in Grades 8, 10 and 12 recorded their classroom observations over a two-year period. During the observations, the researchers were to report the most prevailing activities students were engaged in on a five-minute period. They found that high stakes examination had a great impact on teaching especially in Grade 12. High-stakes examination creates pressure on both students and teachers, and erodes the teachers' ability to creative teaching. Most teachers in Grade 12 reported using the time allotted to teaching science to teach how to write high stakes test.

Orefield and Wald (2000) discussed the unfairness of the system in high stakes test. They reported on minorities and students of low social economic status. They argued that high stakes testing is a way to hold schools accountable for poor and minority students' performance while punishing the students. They noted that the use of high stakes tests as widespread today contradicts the recommendations of institutions such as the National Academy of Science and the Department of Education's Office of Civil Rights regarding the use of the single test on important decisions related to students' achievement. Without rejecting the importance of assessment as "a powerful lever for shaping instruction" (Firestone, Mayrowetz, & Fairman, 1998, .95), as means to measure academic growth, the authors stated that high stakes tests are educationally unsound and appear to discriminate against minorities and students living in poverty. The authors concluded that educators need to find means for holding schools and students accountable for achievement while avoiding penalizing the disadvantaged.

Hoffman, Assaf, and Paris (2001) surveyed a group of teachers in Texas. Participants in the survey were 200 subjects who returned their responses among 500 initially randomly selected. This sample included classroom teachers, reading specialists, curriculum supervisors, and educators in leadership positions, all members of the Texas State Reading Association (TSRA). It is reported that teachers spent 8 to 10 hours of valuable instructions a week for test preparation activities. These include strategies how to do well on the test, motivation to school attendance, teaching or reviewing topics that will appear in the test, test-taking strategies, and having students practice with test forms from previous years. According to respondents, many students experienced headache, stomachache, and other disturbances that might undermine performance on the test, which might in turn adversely impact low-scoring students as well. The authors

concluded that, as implemented, the Texas Assessment of Academic Skills (TAAS) is not only affecting instruction in negative ways, but also is leading teachers and students to drop out.

Educational Outcomes for Students with Disabilities and the School Reform Movement

A number of studies have evaluated special education program outcomes across the states. The main purpose of the investigations was to look at the overall educational and behavioral outcomes of students enrolled in the programs across the nation. The studies have also highlighted the partial failure of the special education program to achieve its intended role. Some of these studies are reviewed below.

Karpinski, Neubert, and Graham (1992) conducted a study of 86 students with mild disabilities (52 had graduated and 34 dropped out of high school) in a predominantly rural school district in a mid-Atlantic State. The students were interviewed at two points in time about their employment, residential status, and participation in postsecondary education and training programs. Information collected was then disaggregated to allow for comparison between the two groups on employment outcomes, participation in postsecondary education and training programs, and residential status. Karpinski, Neubert, and Graham (1992) reported that even though participants in both groups had relatively high rate of employment, the picture concerning participation in postsecondary education was not encouraging. Less than one fourth of the students in the study had participated in a postsecondary program.

The 21st Annual report to Congress on the Implementation of IDEA in 1999 noted that students with disabilities are overrepresented in correctional facilities. The statistics contained in the report are pretty alarming in revealing the negative outcomes of education of youth with disabilities in the nation. The report mentioned that in 1996-1997, 45% of students with LD, 42% of students with Emotional Disturbance, 7% of students with Mental Retardation, 3% of students with Speech or Language impairment, and 3% of Other Disabilities were held in correctional facilities. The students incarcerated might be confined in "jails, detention facilities, group homes for young offenders, adults or juvenile prisons, camps, ranches, private programs or treatment facilities" (p. II-2).

McGrew et al (1994) investigated the achievement outcome information of students across the country. The ultimate goal of the research conducted by the national Center on Educational Outcomes (NCEO) was to gather information that will help produce a policy-relevant report on the educational status of students with and without disabilities. The center has conducted a survey including all the state directors of special education and their designees. Information was gathered on state efforts in areas such as: (a) federally reported data; (b) assessment of outcomes; (c) inclusion of students with disabilities in state assessment; (d) state assessments needs and highlights; (e) activities in selected outcomes areas and practices, programs and plans related to outcomes. The initial survey included 49 of the 50 states of the nation. Of states that reported that some students with disabilities were part of their general education large-scale achievement assessment, 27 or 54% indicated that students with disabilities could be identified in their data sets. Among them, the NCEO was able to secure copies from only six states that represented 22% of the 27 from the previous group and 12% of the 50 states in the country. Large-scale assessment mainly covers reading and mathematics in most states. In other states students are assessed in writing or language. Very seldom is information in subject areas such as social

studies and sciences collected. McGrew et al. also reported that aggregation across states is not always feasible.

Conclusion

The overall purpose of this review was to examine the evidence of the effects of large-scale high stakes assessment on students classified as having mild disabilities. It can be concluded that (a) despite the mandates from IDEA 1997, students with disabilities are still excluded from state and districts accountability system. When they are tested as it appears in most of states, they might be excluded from the accountability system. The estimated prevalence of states with full inclusive accountability system is not always clear to determine. The national data reported by the NCEO on the issue do not always cover the nation, making obtained statistics less reliable. (b) Several conception papers have discussed the negatives impacts that might be associated with full inclusive assessment. Some of the drawbacks are supported by sound research data. For instance, most authors have highlighted narrowing the curriculum and reducing instructional time devoted for subject areas not tested. Wideen, O'Shea, Pye, and Ivany (1997) have used randomly selected group of stakeholders to confirm this assumption even though this particular paper bears a non-negligible limitation. The authors did not specify the demographic characteristics of the population included in the research. For instance we don't know how many students with special needs participate in the study. (c) It appears that participating in high stakes testing does not necessary mean being included in state or district accountability system. When they can be assessed, students with mild disabilities can take the regular test with accommodation, without accommodation, or alternate assessment. Controversies arise when students with disabilities participate in high stakes testing with accommodations, and when they take alternate assessment. Another important finding in this area is that most studies on testing with accommodation focus on students with LD. Again any generalization to the group of students entitled to test accommodation is almost impossible. In addition most papers used intact groups of various stakeholders. Seldom are random samples used. For example statewide reports and surveys provided valuable information. It also can be noted that some of the studies are really explicit in reporting their findings. Fuchs et al. (2000) reported the presence of "boost" but failed to make the reader comprehend this result. High stakes policies have some unintended consequences that might in a long term affect students receiving test accommodations or alternate assessment. For instance it is reported that upon publication of test scores, teachers of low-ranked schools leave the field for better employment, and students whose scores are "flagged" may get low-paid employment. (d) This review of literature on high stakes assessment and students with learning disabilities is enriching. It is philosophically sound to include students with disabilities in states and district accountability system. School officials will take students' educational outcomes more serious. The main lack in the present literature on high stakes assessment is the noticeable absence of parents and students' input. Knowing parents and students perceptions of high stakes assessment and its consequences appears to be very interesting areas to explore. In some states when students pass the test with accommodations, or when they take the alternate assessment, in place of a diploma they receive a certificate. It also is important to examine the meaning of the certificate on the job market and how it might impact the students; social, financial, and emotional well-being.

Return to the Table of Contents

References

American Educational Research Association. (2000). AERA position statement concerning high stakes testing in preK-12 education. Retrieved April 22, 2001 from World Wide Web: http://www.aera.net/about/policy/stakes.htm

Allington, R.L., & McGill-Franzen, A. (1992) Does high stakes testing improve school effectiveness? ERS Spectrum, 10, 3-12.

Allington, R.L., & McGill-Franzen, A. (1993). Flunk 'em or get them classified: The contamination of primary grade accountability data. Educational Researcher, 22, 19-22.

Danforth, S. Rhodes, W. (1997). Deconstructing disability: A philosophy for inclusion. Remedial and Special Education, 18, 357-366.

Ducharme, E.R. & Ducharme, M.K. (1998). To test or nor to test: That is not the question. Journal of Teacher Education, 49, 83-84.

Firestone, W.A. Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. Educational Evaluation and Policy Analysis, 20, 95-113.

Fuchs, L.S., Fuchs, D. Eaton, S.B. Hamlett, C.L., & Karns, K.M. (2000). Supplementing teacher judgments of mathematics test accommodations with objectives data sources. School Psychology Review, 29, 1, 65-85.

Geisinger, K.F. (1994). Psychometric issues in testing students with disabilities. Applied Measurement in Education, 7, 121-140.

Gronna, S.S., Jenskins, A.A., & Chin-Chance, S.A. (1998). The performance of students with disabilities in a norm-reference, statewide standardized testing program. Journal of Learning Disabilities, 31, 482-493.

Gutloff, K. (1999). Is high stakes testing fair? National Education Association Today, 17,

Henley, M. Ramsey, R.S. Algozzine, R.F. (1996). Characteristics and strategies for teaching students with mild disabilities. Needham Heights, MA: Allyn & Bacon.

Hoffman, J.V., Assaf, L.C., & Paris, S.G. (2001). Highstakes testing in reading; Today in Texas, tomorrow? The Reading Teacher, 54, 482-492.

Hollenbeck, K. Tindal, G., Almond, P. (1998). Teachers' knowledge of accommodations as a validity issue in high-stakes testing. The Journal of Special Education, 32, 175-183.

Heubert (2000). High- stakes: Opportunities and risks for students of color, English-Language learners and students with disabilities. Retrieved June 13, 2001 from World Wide Web: http//www.cast.org/ncac/Graduation Promotion Testing 820.cfm

IRA, (1999). High stakes assessment in reading: A position statement of the International Reading Association. Journal of Adolescent and Adult Literacy, 43, 305-312.

Jones, G., Jones, B.D. Hardin, B., Chapman L., Yarbrough, T., & Davis, M. (1999). The impact of high stakes on teachers and students in North Carolina. Phi Delta Kappan, 81, 1999-203.

Kaiser, J.S. (2000). Are high stakes tests taking control? Looking beyond test score to maintain middle level practices into the new millennium. Schools in the Middle, 9(7), 18-21.

Kampfer, S.H. Horvath, L.S., Kleinert, H.L., & Kearns, J.F. (2001). Teachers' perceptions of one state's alternate assessment; Implications for practice and preparations. Exceptional Children, 67, 361-374.

Kleinert, H.L., Kennedy, S., & Kearn, J.F. (1999). The impact of alternate assessments: A satewide teacher survey. The Journal of Special Education, 33, 93-102.

Lacina-Gifford, L.J., & Kher-Durlabhji, N. (1993). The price of high stakes testing on children. Education, 112, 565-566.

Landau, J.K., Vohs, J.R., & Romano, C.A. (1998). All kids count. Boston, MA: Federation for Children with Special Needs.

Lagenfeld, K., Thurlow, M.L., & Scott, D. (1997). High stakes testing for students: Unanswered questions and implications for students with disabilities. Minneapolis, MN: National Center on Educational Outcomes.

McDonald, L.R., & Bean, L.C. (1998). Thinking of retaining a student? Try one or more of the twenty-five alternatives to retention. Education, 112, 567-570.

McGrew, K.S., Spiegel, A.N., Thurlow, M.L. Shriner, J., & Ysseldyke, J.E. (1994). Secondary analysis of state assessment data: Why we can't say much about students with disabilities. Minneapolis, MN: National Center on Educational Outcomes.

McGrew, K.S., Thurlow, M.L., Spiegel, A.N. (1993). An investigation of the exclusion of students with disabilities in national data collection programs. Educational Evaluation and Policy Analysis, 15, 39-352.

National Council on Disability (1993) Serving the nation's students with disabilities: Progress and Prospects. Washington, DC. Retrieved June 30, 2001 from World Wide Web: www.ncd.gor/newsroom/publications/progress.html

Orfield, G., & Wald J. (2000). The high stakes testing manis hurts poor and minority students the most: Testing, testing. The Nation, 270, 38-48.

O'Neill, P.T. (2000). Pass the test or no diploma: High stakes graduation testing and children with learning disabilities. LDOnLine. Retreived April 25, 2001 from the World Wide Web: http//www.ldonline.org/ld_indepth/assessment/oneill.html

Pipho, C. (200). Stateline- The sting of high stakes testing and accountability. Phi Delta Kappan, 81, 645-651.

Phillips, S.E. (1994). High stakes testing accommodations: Validity versus disabled rights. Applied Measurement in Education, 7, 93-120.

Popham, W.J. (1987). The merits of measurement-driven instruction. Phi Delta Kappan, 68, 679-682.

Ransom, K.A., Santa, C.M., Williams, C.K., Farstrup, A.E., Kathrny, H., Baker, B. M., Edwards, P.A., Hoffman, J.V., Klein, A.F., Larson, D.L, Logan, J.W., Morrow, L.M., & Shanahan, T. (1999). High stakes assessment in reading: A position statement of the International Reading Association. Journal of Adolescent and Adult Literacy, 43, 305-312.

Stake, R. (1998). Some comments on assessment in U.S. Education. Education Policy Analysis Archives, 6 (14). Retrieved on April 20, 2001 from World Wide Web: http://www.olam.ed.asu.edu/epaa/v6n14.html

Stiggins, R.J. (1999). Assessment, student confidence, and school success. Phi Delta Kappan, 81, 191-198.

Stecher, B.M., & Barron, S.I. (1999, June). Quadrennial milepost accountability testing in Kentucky. CSE Technical Report 505.

Tingley, S. (1999). Weighing the cattle. Education Week, 18 (43), 44-44.

Thompson, M.L. (2001, April 11). Use of accommodations in state assessments-what databases tell us about differential levels of use and how to document the use of accommodation. Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle Washington.

Wideen, M.F., O'Shea, T., Pye, I., & Ivany, G. (1997). High stakes testing and the teaching of science. Canadian Journal of Education, 22, 428-444.

Return to the Table of Contents

To top