

5-1-2005

High-stakes Testing and Special Populations

Gary H. Sherwin

California State University San Bernardino

Todd Jennings

California State University San Bernardino

Follow this and additional works at: <https://scholarworks.lib.csusb.edu/wie>



Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), [Curriculum and Instruction Commons](#), [Disability and Equity in Education Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), [Educational Psychology Commons](#), [Elementary Education and Teaching Commons](#), [Junior High, Intermediate, Middle School Education and Teaching Commons](#), [Secondary Education and Teaching Commons](#), and the [Social and Philosophical Foundations of Education Commons](#)

Recommended Citation

Sherwin, Gary H. and Jennings, Todd (2005) "High-stakes Testing and Special Populations," *Wisdom in Education*: Vol. 1 : Iss. 1 , Article 3.

Available at: <https://scholarworks.lib.csusb.edu/wie/vol1/iss1/3>

This Article is brought to you for free and open access by CSUSB ScholarWorks. It has been accepted for inclusion in *Wisdom in Education* by an authorized editor of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.

High-stakes Testing and Special Populations

Abstract

This opinion paper critically examines the use of high-stakes testing on special populations. Without appropriate accommodations, standardized exams are not valid for some students with special needs. Unfortunately, many classroom teachers who must initiate testing accommodations lack knowledge of appropriate accommodations and regularly fail to provide the necessary testing accommodations. The deficit understanding of testing accommodations makes comparisons between classrooms, schools, and districts invalid since some scores lose validity. Solutions specific to standardized testing and students with special needs are offered and a more encompassing solution to the problems incurred from these tests when used for high-stakes is suggested.

Keywords

testing, curriculum, standardized tests, high stakes tests

Author Statement

Dr Gary H Sherwin and Dr Todd Jennings are professors in the College of Education, California State University San Bernardino

High-stakes Testing and Special Populations

Gary H. Sherwin and Todd Jennings

Department of Educational Psychology and Counseling
California State University San Bernardino

Abstract

This opinion paper critically examines the use of high-stakes testing on special populations. Without appropriate accommodations, standardized exams are not valid for some students with special needs. Unfortunately, many classroom teachers who must initiate testing accommodations lack knowledge of appropriate accommodations and regularly fail to provide the necessary testing accommodations. The deficit understanding of testing accommodations makes comparisons between classrooms, schools, and districts invalid since some scores lose validity. Solutions specific to standardized testing and students with special needs are offered and a more encompassing solution to the problems incurred from these tests when used for high-stakes is suggested.

High-stakes Testing and Special Populations

In a decidedly unscholarly vein, this paper will begin and end with fictitious stories. They are fiction because the authors created the stories. In the first two stories, it is possible that similar real-life stories are being enacted on school campuses across the nation. The reader will determine the degree of fact or fiction given the current pervasive culture of high-stakes testing in education. This paper is designed to be an introduction to some of the issues regarding standardized assessment and students with special needs. Therefore, the reader does not need to know about criterion related validity or split-half reliability.

Early on the distinction between standardized tests and high-stakes testing must be established since the use of a standardized test is potentially problematic for a number of reasons beyond the scope of this article, and a standardized test in and of itself is not high-stakes testing. However when the data generated by standardized tests are used to determine or allocate opportunities such as determining who graduates, which teachers get raises, or which administrators retain their jobs, the results from the standardized test become high-stakes testing. Therefore, high-stakes testing will be seen as the use of scores on standardized academic achievement or performance tests to make decisions that have far-reaching consequences (both personal and political) for examinees under the assumption that such efforts promote accountability (Darling-Hammond, 2002).

For the purposes of this paper, children and youth with special needs will be operationally defined as students with one or more pervasive psychological, processing, and/or physical disorder(s) that manifests across environments and has an adverse effect on school performance. The condition(s) must adversely affect one or more major life activities and it is believed the student will benefit from special education services. Some students may fit this description and not be identified for special education services. Following are two “fictional” case studies.

Fictional Case Studies

John

November 11, 2005

Dear Teacher,

Ya, so I am suppose to write this stupid interactive journal. You said to be honest. I bet you don't mean it and probably don't even read these lame things. Anyway that big test is next week. If I don't pass it, I guess I don't graduate. My buds say only retards don't pass the test. Retards don't get to party. I never passed one of those bogus tests in my life. I guess that makes me a retard, duh! But not really because I figured out a way to mess with the test. I just mark any answers on purpose. That way, I can't be called a retard because I screwed with the test. My buds can't believe I don't care. I guess they don't know what I really care about. It's going to be funny even if I get in trouble.

John

John is a high school senior who has been in special education for four years. After seven years of school failure, a teacher finally noticed John learned differently and marched through the labyrinth of student study teams and testing to eventually have an IEP team agree that John was learning disabled. Following a couple years of inexperienced and un-credentialed special education teachers, John has had some talented and trained teachers - with proper accommodations, he may have passed the test if he tried. After seven years of failure, he really couldn't chance trying and possibly failing. He couldn't be the dummy once again. Never mind that he will receive a Certificate of Completion rather than graduate. Never mind that the teacher's expertise will be called into question and that if enough students react as John did, administrators' and educators' careers can be ruined. John did in fact get invited to all the graduation parties. The effects of high-stakes testing and his Certificate of Completion will remain with John long after the sweet memories of the parties fade.

Chris

On another high school campus in another district, Chris is excited about taking the big test. She was relieved when she was given special education services in her general education classroom because she finally understood why it was so difficult for her to read. Chris has a visual processing disorder - she is dyslexic. After intensive reading training and accommodations such as extended time to take examinations, her grades of 'D's shifted to mostly 'B's. Given the gift of time, the tests actually measured what Chris knew rather than the degree dyslexia affected her reading. Chris was ready for the big test and excited about graduating. However, Chris didn't graduate. Her district interpreted the testing guidelines in No Child Left Behind (NCLB) to mean that Chris could not have extended time to take the test. Despite her Individualized Education Plan (IEP), a legally binding document that called for time and half to complete all tests, she was not given her rightful accommodation. Consequently, the State competency exam simply measured the extent to which dyslexia affected her ability to read rather than her comprehension of the material. Once again, high-stakes testing drastically changed Chris' life because she didn't graduate and the exam results could

have a negative affect on her school administrators and teachers as well. As an experiment, Chris was given time and a half to complete the test. She passed and was in the top half of the testing pool. Of course these scores couldn't count since the test was "altered" given the District's interpretation of NCLB guidelines. Forget that Chris had proven she knew the material—one inflexible high-stakes test had shown she did not deserve to graduate and that her teachers were in fact inferior.

History

While this paper could examine the validity of a single test to determine major outcomes—that would be old wine in new bottles for most educators, including those in special education. Special education requires that all students receiving special services under Individuals With Disabilities Education Act (IDEA 04) are categorized and labeled with a handicapping condition such as emotionally disturbed or learning disabled. High-stakes testing in special education could mean that a student is labeled on the basis of a single standardized assessment. However, court cases such as *Larry P. vs. Riles* (1972) determined that the use of use of Intelligent Quotient (IQ) scores as the sole measure to identify African American students for special education services is illegal since the tests were found to be biased. Subsequent legislation (IDEA 97; IDEA 04) now forbids any single standardized exam being used to make major decisions regarding placement or identification of a student with special needs. Rather special education law requires a testing battery that includes standardized and authentic tests performed by a multidisciplinary team in the student's native language (IDEA 04). Unfortunately, in the current climate of educational reform it appears that the authors of NCLB have ignored the hard lessons that were learned in special education during the past 30 years.

To continue, this paper will examine the strengths/limitations and needs of children and youth with special needs regarding standardized examinations. Despite current trends found in education and laws crafted by politicians such as NCLB, a critical examination of the unintended outcomes on different populations resulting from the dogmatic application of high-stakes testing is needed because trends and laws do not necessarily translate into best practice.

Special Education Participation in High-stakes Testing

Beginning with the Individuals with Disabilities Education Act Amendments of 1997 (IDEA 97) and strengthened through the Individuals with Disabilities Education Improvement Act of 2004 (IDEA 04), it is now expected that children and youth with special needs are included in statewide assessments. Students with special needs were exempted from statewide testing until the 1990's. There were two primary reasons to include students with special needs in statewide assessments. First, the movement to include special need students in statewide assessments was in part an effort to thwart districts improving test scores by over-identifying students in special education since students categorized with special needs were exempt from statewide testing. Multiple studies during the 1990's showed that passing rates on state competency tests dramatically improved when more

students were placed in special education and were thus exempt from statewide testing since they were essentially removed from the testing pool (Schulte, Villwock, Whichard, & Stallings, 2001). For example, Allington and McGill-Franzen (1992) found that low-performing schools in New York showed vast improvement in State competency exam pass rates when they increased the percentage of students in special education. Those previously low-performing schools improved pass rates by increasing the percentage of special education students to be three times greater than the rate of special needs students in schools that were not low-performing.

The second reason to include children with special needs in statewide testing is found in the common goal to successfully include special education students in general education classrooms. At this time, students with learning disabilities spend the bulk of their day in general education classrooms and are a majority of the special education population (US Department of Education, 2000). Since high-stakes testing has become a routine organizing element of general education classrooms, special education students are swept up in the high-stakes testing “net” along with their non-special education peers. This occurs despite their unique needs as a result of exceptional conditions that require appropriate and consistent accommodations in order to achieve valid test results. Unfortunately, teachers may not be equipped to determine testing accommodations and the gray area between testing accommodations and testing modifications may eliminate the use of some important accommodations.

Standardized Testing and Special Education

Testing accommodations and modifications

Special education law (IDEA 97) mandates testing accommodations when needed to ensure that test results are valid while allowing for a small percentage of students to complete a modified or alternative assessment. When an alternative assessment is administered in lieu of the statewide graduation exam, the results are disaggregated and commonly lead to a Certificate of Completion rather than graduation with a diploma. Whether the test is modified or accommodations are designed for specific students, these changes must be specified in the students’ Individual Education Plan (IEP). Following is a brief discussion to help distinguish between testing accommodations and testing modifications.

Testing accommodations work to explicitly maintain the construct(s) being testing while altering access to the test or changing the means to respond to the questions by the test taker. In this way, exceptional students may complete the tasks within the test without confounding influences of format, administration, or response type. For example, testing accommodations include changing the format to Braille, altering administration if extra time is given to complete the exam, and individualized response if a student who has great difficulties writing is allowed to respond to the test questions using a large keyboard computer. Therefore, the accommodations are unique and specific to each student. Differing from testing accommodations, *testing modifications* alter the test constructs measured and results in equal effects for all test takers (Hollenbeck, Tindal, & Almond,

1998). As an example to clarify the distinction between *accommodation* and *modification*, if a standard written test were administered to a student who is blind and reads Braille, the resulting answers would not measure the student's knowledge of the constructs but would only measure the degree to which the student can "read" non-Braille text. Therefore, this would be a testing *accommodation* because the only thing changed would be to administer the test in Braille to the student. A test *modification* would have occurred if the test administrator decided to remove all questions that had cue words involving visual concepts such as "brighter or lighter" since these words may have different meanings for the student who is blind. This modification will change some of the constructs measured and the changes will be applied to all students taking the test. Students who are severely cognitively impaired make up the majority of students who receive modified assessments (Hollenbeck, Tindal, & Almond, 1998). Modified tests are administered because even with germane testing accommodations, a population of students may not have learned or been systematically taught many of the complex constructs that are assessed in the state competency exams due to their exceptional needs. Rather than frustrate students by testing them on unfamiliar constructs, the alternative statewide competency exam addresses the statewide curricular standards using constructs familiar to the students. For example, geometry might be approached via real life experiences such as a stop sign.

Valid test results are readily achievable when appropriate testing accommodations are applied to applicable students with special needs. The objective of testing accommodations is to "level the playing field" and remove potential barriers that are unrelated to the test content so that all test takers have equal access to the questions and means to express answers. The definition of equality in this case is not that everyone gets the same testing conditions. Rather, equality means that everyone has full access to the opportunity to demonstrate what he or she knows.

Chris, the second case study, has a visual processing disorder and requires extra time to process written information. If it takes the average student two minutes to process a paragraph, it will take Chris three minutes to process the same paragraph. Therefore, the accommodation of additional time to complete the exam simply removes a barrier to answering the question but does not alter what Chris is expected to know. Testing accommodations must be specific to the student since additional time for the student who is blind and taking the test in Braille would give him an unfair advantage assuming she or he processes Braille at the same rate most students read print. When appropriate and individualized testing accommodations are used, the resulting scores have greater validity since they are a better representation of what the student actually knows since access to the questions are not hindered by physical, emotional, or cognitive functioning differences.

The teachers' role in testing accommodations.

Under NCLB (2001), students with special needs are considered to be a disaggregated group and the law requires a minimum 95% participation rate for each of the disaggregated groups. A student is deemed to have participated in the general assessment only if his or her test score is counted in the statewide accountability system. That means that if the testing accommodations provided to a

student with a disability results in the scores being discredited as unreliable or invalid, the scores of that student will not count as part of the 95% required participation rate. Most States have devised guidelines for accommodations that do not render the State Achievement Exams unreliable or invalid. This brings up three questions;

1. Do teachers understand the acceptable testing accommodations?
2. Do educators grasp which students should be identified for accommodations while matching optimal testing accommodations to individual students?
3. Do the acceptable testing accommodations omit accommodations that will adversely affect testing validity and reliability?

The decision to pursue accommodations and initial recommendations for accommodations begins with the classroom teacher. Unfortunately, there is great variability in teachers' decisions regarding appropriate accommodations. This renders valid comparisons between classrooms, schools and districts virtually impossible. These inconsistencies in employing accommodations confound the uniformity needed to ensure the validity of standardized assessments and the meaning of comparative test data. Hollenbeck, Tindal, and Almond (1998) found that overall, only 21% of teachers in a survey of 52 general education teachers and 62 special education teachers used the accommodations specified in the testing manual and they only understood 55% of the total allowable accommodations. Therefore, less than one quarter of the teachers used testing accommodations and they lacked knowledge of nearly half of the available accommodations. To illustrate, consider the following example. Classroom teacher A identified the appropriate students and employed the best accommodations which lead to exam scores that showed appropriate academic growth for all his/her students. Classroom teacher B, due to a lack of knowledge, did not apply any accommodations to his/her students. As a result, 11% of his/her students displayed little or no significant academic gains on the statewide exams despite the fact that both groups of students learned all of the same material to the same degree. If teacher B had used accommodations, his/her students would show growth equal to teacher A. As a result, the test results for teacher B did not measure achievement but rather the disabling conditions that impeded access to the test. Teacher B created a testing situation which rendered some of the testing scores invalid therefore making comparisons between classrooms meaningless. Obviously, classroom teachers need knowledge about assessment and accommodations if high-stakes testing is to have any validity. Unfortunately, assessment instruction is not included in many of the teacher education

programs. Data shows that nearly half of teacher education programs in the United States do not require a course in measurement for initial certification (Schafer, 1991; Stiggins, 1991). NCLB allows individual states to determine which changes to the statewide exams are accommodations and which changes are modifications. The latter potentially confounds test score validity and is thus disallowed. However, some testing accommodations that individual states eliminate could adversely affect the reliability and validity of statewide assessment scores as well. Thurlow, Lazarus, Thompson, & Morse, (2005) find that the three accommodations states are most prone to eliminate from statewide testing are reading the test out-loud, the use of scribes, and using a calculator. Granted, depending upon the constructs being measured by the testing instrument, all three of these accommodations could be seen as a test modification rather than an accommodation. However, for a student with a severe visual processing deficit, having a test monitor read the questions out-loud may be the only means for him or her to access the test. A scribe may be the sole means that a student with severe limb spasticity can communicate the test answers. If the test is measuring mathematical algorithms, then a calculator would be a testing modification. However a calculator may be the best means to determine if some students understand the correct mathematical process to answer the question.

Conclusions

In conclusion, without appropriate accommodations, some students with special needs will exhibit scores that reflect his or her handicapping condition rather than his or her knowledge of the constructs on a standardized exam. When that happens, the standardized scores become invalid and legitimate comparisons between students, classrooms, schools, and districts are impossible. Unfortunately, the classroom teachers who initiate testing accommodations may not be prepared to do so since many teacher preparation programs do not include rigorous curriculum on assessment constructs and procedures nor even critical examinations of standardized testing procedures. While the soundness of high-stakes testing must be examined as appropriate for any students, the following are some suggestions to improve standardized testing relative to special education students specifically.

- Do not rely on one written test. Perform a test battery that includes authentic and standardized assessments performed by a multidisciplinary team in the student's native language.
- Ensure that all students who need testing accommodations receive the appropriate accommodations noting that there are students with special needs that may not yet be identified.
- Campaign to include comprehensive curriculum about assessment in teacher preparation programs and provide in-services that informs teachers about available testing accommodations and selecting applicable students for accommodations.

- Bring educators and the public to an improved understanding of the strengths and limitations of standardized assessments, thus producing actual accountability.

The Concluding Fictions

In the final analysis, this paper only offers a small tattered bandage for the larger festering wound of high-stakes testing. However, since the paper began with two fictions, it will also end with two fictions. Once again, the readers will gauge truthfulness, but this time, the readers will also determine the final outcomes. In this fiction, the authors realize that they failed to ask the important questions. The readers accept two underlying assumptions;

1. that a single standardized test is inadequate to measure a significant and comprehensive body of knowledge and may be invalid for distinct populations, and,
2. the brightest future will belong to the citizenry that understands how to quickly access information and is able to sort fact from fiction in the process of decision making based on data found on unmediated forums such as the Internet.

In this story, the authors and readers accept the underlying assumptions, so the two primary questions that should be asked are;

1. Why have professionals in education allowed people who do not understand the limitations of standardized testing to dictate high-stakes testing in the first place?
This has resulted in an attempt to impose invalid accountability on the teaching workforce who is capable of self-monitoring and,
2. Why have thinking people remained mute when educational careers and young futures are controlled by tests that require the regurgitation of superfluous information evaluated under questionable circumstances?

The information measured in most statewide standardized testing is of dubious worth since that body of knowledge is quickly and readily accessible via sources such as the internet or hand-held devices ranging from calculators through small computers. This makes the ability to access, utilize, manipulate, incorporate and evaluate all this information more important than simply recalling or recognizing the information on a test. Furthermore, standardized exams commonly ignore the constructs most would agree are needed to succeed such as the ability to access the applicable data and then reflectively and creatively make decisions that are more complex than forced multiple choice options. Accessing the cornucopia of information via technology requires discrete skills coupled with expert judgment in potentially unmediated mediums awash with both dubious and sound information. Perhaps if we eliminate the misuse of standardized testing as high-stakes testing, education can then take a critical look at the demands of a new century and adjust curriculum to improve educational outcomes rather than constantly reacting to ignorant and misguided attempts to save education by repeating past practices which privilege assumptions of homogeneity while punishing diversity. It is now up to the reader to determine if that will be fact or fiction. Since this introductory article did not delve into specific testing accommodations or myriad issues of standardized testing being used as high-stakes testing, following are additional resources for the interested reader:

Additional Resources

General high-stakes/standardized testing

- Alfie Kohn Organization URL: <http://www.alfiekohn.org/>
- No Child Left Behind.com URL: <http://www.nochildleft.com/>
- The National Center for Fair & Open Testing URL:
<http://www.fairtest.org/>
- California Department of Education (NCLB) URL:
<http://www.cde.ca.gov/nclb/index.asp>
- U.S. Department of Education (NCLB) URL:
<http://www.ed.gov/nclb/landing.jhtml?src=pb>

Standardized testing and special education

- LD On-Line (Aligning Special Education with NCLB) URL:
http://www.ldonline.org/ld_indepth/special_education/alignment_primer.html
- CEC (NCLB Implications for Special Education Policy and Practice)
URL: <http://www.cec.sped.org/pp/NCLBside-by-side.pdf>
- California Department of Education (Special Education) URL:
<http://www.cde.ca.gov/sp/se/>

References

- Allington, R. L. & McGill-Franzen, A. (1992). Unintended effects of educational reform in New York [Electronic version]. *Educational Policy*, 6, 397-414.
- Darling-Hammond, L. (2002, June). Standardized testing: What's at stake in high-stakes testing [Electronic version]?. *The Brown University Child and Adolescent Behavior Letter*. 18(1), 1-3.
- Hollenbeck, K., Tindal, G. & Almond, P. (1998). Teachers' knowledge of accommodations as a validity issue in high-stakes testing [Electronic version]. *The Journal of Special Education*, 32 (3), 175-183.
- Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-466, §20 U.S.C. 1400 et seq. (2004).
- Individuals with Disabilities Education Act Amendments of 1997, Pub. L. No. 105-17, §20 U.S.C. 1400 et seq. (1997).
- Larry P. v. Riles. C-712270-RFP (N.D. Ca. 1972), 495 F. Supp. 96 (N.D. Cal. 1979) Aff'r (9th Cir. 1984), 1983-84 EHLR DEC. 555:304.
- No Child Left Behind Act of 2001, PL 107-110, 20 U.S.C. §§ 6301-6578 et seq.
- Schafer, W.D. (1991). Essential assessment skills in professional education of teachers [Electronic version]. *Educational Measurement: Issues and Practice*, 10(1), 3-6.
- Schulte, A.C., Villwock, D. N., Whichard, S. M. & Stallings, C. F. (2001). High-stakes testing and expected progress standards for students with learning disabilities: A five-year study of one district [Electronic Version]. *School Psychology Review*, 30 (4), 487-506.

Stiggins, R.J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice*, 10(1), 7-12.

Thurlow, M.L., Lazarus, S.S., Thompson, S.J., & Morse, A.B. (2005). State policies on assessment participation and accommodations for students with disabilities. *The Journal of Special Education* 38(4) 232-240.

U.S. Department of Education, (2000). *Twenty-second annual report to Congress on the implementation of the Individuals with Disabilities Education Act*. Washington, DC: Author.