

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to the *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited.

Volume 13, Number 6, June 2008

ISSN 1531-7714

Unexpected Testing Practices Affecting English Language Learners and Students with Disabilities under No Child Left Behind

Mark Fetler

The testing and accountability requirements of the No Child Left Behind Act impose sanctions on schools for not making adequate yearly progress in student achievement. The sanctions may encourage inappropriate practices intended to raise scores of low performing student subgroups. This article considers evidence and consequences of misclassification of English language learners as students with disabilities.

The Commission on No Child Left Behind (2007) describes the poor achievement of English language learners (ELLs) and students with disabilities (SWDs) as alarming. Failure was not expected from a law that intended to improve teaching and learning for students in poverty or otherwise at risk. The No Child Left Behind Act (NCLB) defines educational equity to include not only access to instruction, but also testing and accountability for all, specifically embracing ELLs and SWDs.

ELLs and SWDs constitute a significant portion of public school enrollment. (Hoffman and Sable, 2006) Nationally, in 2003-04 there were 10.6 percent ELLs and 13.6 percent SWDs. While these subgroups are a minority of the total population, they are a majority of the students targeted by NCLB. The SWD and ELL subgroups intersect and students who belong to both possess complex needs and legal protections. Although some ELLs are correctly identified as having cognitive or other disabilities, limited English proficiency (LEP) status is not by itself considered to be a disability.

This article examines evidence of the misclassification of ELLs as SWDs and the unexpected consequences of this error. In many states the numbers of ELLs are too small to permit a useful analysis of the overlap between ELLs and SWDs. However, in a few large states and nationally such analyses are feasible and may usefully inform education practice and policy. California, a large state where over a quarter of enrolled students receive ELL services, is conducive to an examination of the intersection between special education and ELL programs.

BACKGROUND

Title I of NCLB requires academic standards and testing, aligned with those standards for all students, including ELLs, with reasonable accommodations if needed, to the extent practicable in the language most likely to yield accurate and reliable scores. Title III of NCLB requires states to adopt English language proficiency (ELP) standards to guide learning English as a second language, and to administer to all ELLs a technically sound English language proficiency test

aligned to those standards. Testing results under both Titles of NCLB feed into accountability systems that may either reward or punish schools and districts. The test scores may also influence an individual's grades, promotion, ELL status, SWD status, and funding for services.

The Theory of NCLB

The theory underlying NCLB posits that improved learning results from explicit standards for what students should learn, a schedule for meeting learning targets, tests that measure progress toward the targets, and incentives linked to success or failure. (Linn, 2005) Tests measure learning and guide instruction, and incentives motivate students, teachers, administrators, and parents to work harder. Schools are to make progress toward the targets overall, as well as for racial/ethnic groups, poor students, ELLs, and SWDs. The consequences for schools and districts of failure may include reduced flexibility, increased oversight, and eventual take-over.

The theory omits consideration of the unintended effects of sanctions on parents, students, teachers, and administrators. (Koretz, 2005) Test scores can be manipulated in ways that are unrelated to teaching and learning of the content in the standards. Teaching to the test, exploitation of loopholes in testing procedures, or exclusion of poor performers can raise scores. Without rigorous evaluation of the results educators and policy makers can have little confidence that the theory of NCLB describes reality.

Testing Special Populations

Abedi (2005, 2007) asserts that correct identification of English learners is needed to fairly implement the academic tests imposed by Title I of NCLB. The ELP test scores required by Title III are a logical basis for classifying ELLs. These tests measure language skills, not academic achievement. In practice, less appropriate measures are sometime used, including achievement test scores, immigrant status, number of years in the United States, teacher evaluation, and parent opinion. The use of these other measures to make classification decisions varies widely within and across states. Differences in the measures result in diverse, possibly unsound, often incompatible, definitions of "ELL" across states, districts, and schools.

The Individuals with Disabilities Education Act of 2004 (IDEA) requires SWDs to take statewide

tests, with reasonable accommodations, if necessary. Pullin (2005) NCLB has similar requirements and directs states to combine the scores of SWDs with the scores of all other students and to summarize them separately. A team that includes educators and parents evaluates disability status and creates an Individualized Education Program (IEP) for each student's instruction and testing, based on specific needs. Ideally each state sets forth general procedures for accommodations, and the local IEP team makes informed decisions, depending on the specific needs of the individual student and the characteristics of the test. If the student has a severe cognitive disability, for example a serious neurological defect that significantly hinders cognition, the IEP team may decide to provide the student with an alternate assessment that better suits his or her abilities. Alternate assessments must align with standards, provide valid and reliable results, and be included in the accountability calculations.

Some ELLs are appropriately classified as disabled. However, when language acquisition problems persist and IEP teams lack the linguistic expertise needed to make accurate diagnoses, LEP status may masquerade as a cognitive defect. Teams may believe that providing ELLs with additional services and testing accommodations, that are only available to SWDs, can raise test scores. It is ironic that misclassification of ELLs as cognitively impaired, might track them into instructional programs that do not address their language learning needs, and worsen rather than improve educational equity. Misplacement in special education can be an unintended consequence of testing and accountability programs. Roderick, Nagaoka, and Allensworth's (2005) evaluation of Chicago's use of achievement tests to promote or retain students identifies a case in point. They found little change in teaching and achievement. However, retained students often ended up in special education, where the testing requirement no longer applied.

METHODOLOGY

Data

English Language Proficiency Tests

The California Department of Education (CDE) oversees the California English Language Development Test (CELDT), aligned with the state's ELP standards, pursuant to NCLB Title III English

language proficiency testing requirements. (2007a) School districts administer the CELDT annually to enrolled ELLs during a testing window that runs from July through October. New students who are identified on the basis of a home language survey take the CELDT within 30 days of enrollment. Districts may administer the test to new kindergarten students in July and August, prior to actual attendance at a school.

Students receive a form of the CELDT for the grade-cluster (K-2, 3-5, 6-8, or 9-12) that reflects their placement. The test assesses reading, writing, speaking, and listening skills, except that the kindergarten and grade one form only assesses speaking and listening. Reports include scale scores and proficiency levels (Advanced, Early Advanced, Intermediate, Early Intermediate, and Basic). The total scores analyzed below combine the equally weighted component scores. The 2006-07 edition of the test implements a common scale that allows comparisons of scores over years and across grade clusters.

NCLB's alternate assessment requirements apply to Title III English proficiency testing. (Tomalis, 2003). ELLs who cannot meaningfully take the CELDT, even with an accommodation, receive the lowest possible score on the test, and take an alternate test. California supplies a list of commonly available tests (for example, the "Alternative Language Proficiency Instrument," the "Basic Inventory of Natural Language," or the "Student Oral Language Observation Matrix"). Results are to be used locally, but school districts must send in completed CELDT student information forms for all examinees, including those who take an alternate assessment.

Academic Achievement Tests

CDE (2007b) oversees the California Standards Tests (CSTs) to satisfy NCLB Title I academic testing requirements. In the spring of each year students in grades two through eleven take CSTs in English language arts (CST-ELA) and mathematics. Each grade-level form of the test is aligned to California's content standards appropriate for that grade. The tests are not equated across grades and scores reported at one grade cannot be compared with those for other grades.

All SWDs take the tests with accommodations or modifications, as directed by their IEP team. Students

with significant cognitive disabilities who cannot take the CSTs, even with accommodations, take the California Alternate Performance Assessment (CAPA), which assesses how well they have achieved selected content standards in English language arts and mathematics.

All English learners in grades two through eleven are to take either the CSTs or the CAPA in English. Allowable accommodations for English learners include using English to primary language translation glossaries and translated test directions. CDE reports CST and CAPA results as scale scores and performance levels (Advanced, Proficient, Basic, Below Basic, Far Below Basic).

The National Assessment of Educational Progress (NAEP) is a federal program that periodically conducts assessments in academic subjects. (NCES, 2008) Representative samples of schools and students participated in the 2007 reading assessment. Reports display the results as scale scores and as achievement levels (Basic, Proficient, Advanced). Testing accommodations are to be provided for SWDs and ELLs, who need altered procedures to fairly demonstrate their abilities. NAEP encourages inclusion of SWDs and ELLs if those students participated in the regular state academic assessment and they can participate with accommodations allowed by NAEP. Whether and how special needs students participate in NAEP vary in accordance with states' policies and their implementation by local school and district staff.

Data Analyses

The state level results analyzed below were collected during the 2006-07 school year. CDE's public data website (2007c) furnished all data reported in this article. Total enrollment came from the California Basic Educational Data System (CBEDS) survey conducted in October, and SWD enrollment came from the December survey of special education programs. Some participation rates are estimated using counts from different surveys that were administered at different times and include some error attributable to the migration of students over time. However, at the state level this error should be acceptable.

The CDE does not report CST results broken down by ELL and SWD factors combined. Fortunately, the NAEP Data Explorer provides a

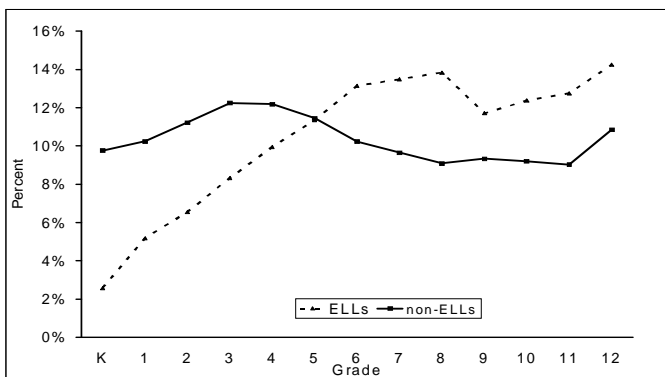
similar national breakdown of NAEP 2007 reading scores. Because NAEP defers to state and local practice for inclusion of ELLs and SWDs (in contrast with California’s universal participation requirement), the NAEP results cannot be directly generalized to California. However, the national results do provide a reasonable basis for comparing scores of these subgroups.

RESULTS

Participation in Special Education

Figure 1 exhibits participation by grade in special education of ELLs and non-ELLs. Students with severe cognitive disabilities who cannot meaningfully take the test even with accommodations are excluded from these subgroups. The dotted line displays the percentage of ELLs who are students with disabilities and is obtained by dividing the total number of CELDT takers into the number with IEPs. Across all grades 1,709,085 students took the CELDT in 2006-07, of which 157,753 or 9.2 percent were SWDs. The trend rises over grades, with the exception of a grade 9 dip, possibly related to dropout, retention, or inefficient transfer of pupil records to high schools.

Figure 1: Estimated Participation in Special Education



Total non-ELL enrollment was estimated by subtracting the total number of CELDT examinees from the total statewide enrollment of 6,279,929 students. Enrollment of non-ELLs in special education was estimated by subtracting the number of CELDT examinees with IEPs from the total special education enrollment of 625,982 students, yielding 10.2 percent of non-ELLs who are SWDs. The corresponding grade-level percentages produce a trend line for non-ELLs that is flat in comparison with the ELL line, displaying a bulge at grades 2, 3 and 4, and a trough after the fourth grade. Because the

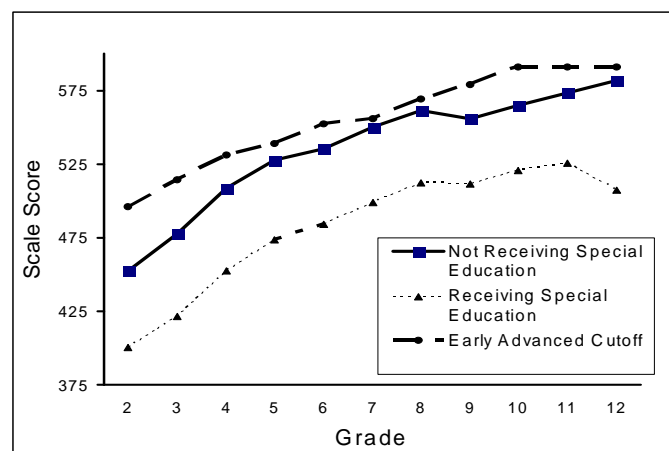
underlying enrollments are large, the standard errors of the percentages are small, ranging from .1 to .2 percent, and the differences between the lines are statistically significant ($p < .001$) at all grades, except at grade 5, where they cross.

The crossing lines in Figure 1 suggest that the decision to classify an ELL as a SWD interacts with grade-level. Before the fifth grade SWDs are underrepresented among ELLs, and later they are overrepresented. The low kindergarten percentage may stem from the fact that the CELDT is administered at the beginning of the school year before many are evaluated for disabilities.

English Language Proficiency

Figure 2 displays average total CELDT scores for SWDs and not-SWDs in grades two through twelve in which all four components of the test are administered. The scores are on a common scale that extends across grade levels. To put the scores in context the dashed line with solid circles displays the CELDT lower limits for the Early Advanced (EA) proficiency level. The CELDT proficiency standards state that students performing at the EA level begin to combine the elements of the English language in complex, cognitively demanding situations and are able to use English as a means for learning in content areas.

Figure 2: California English Language Development Test Results



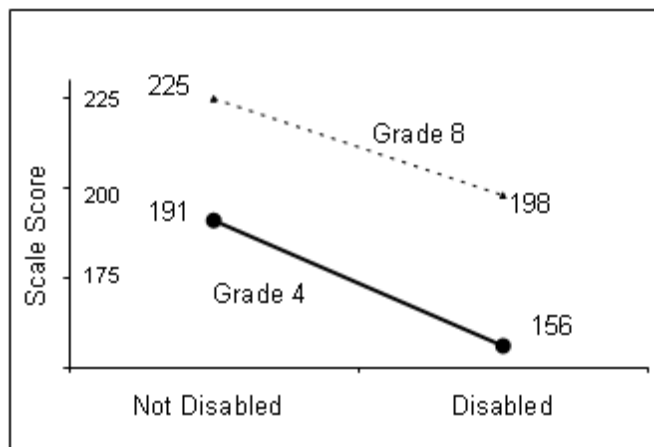
California’s guidelines recommend that a new student scoring below EA be classified as an ELL. ELLs that score at least at the EA proficiency level on the CELDT, at the Basic performance level or above on the CST-ELA, and meet other criteria for

academic success are to be considered for possible reclassification to fluent English proficient (FEP) status. It is more difficult to clear the CST-ELA hurdle than the CELDT criterion. To illustrate, in the eleventh grade in 2007, 21 percent of ELLs scored Basic or better on the CST-ELA, compared to 41 percent scoring EA or better on the CELDT.

Both lines trend upwards in tandem across grades. Non-SWDs, on average, score ten to fifteen points below the EA cut-point. SWDs score about 50 points lower than non-SWDs at all grades, except grade twelve where there is a larger gap. The standard deviation of the overall CELDT scores is approximately 50 points, (CDE, 2007d) yielding a large effect size of about 1.0 for each grade level. A gap of this magnitude that persists across grades suggests that few special education ELLs will reach the minimum CELDT score required for consideration to be reclassified.

California does not report CST-ELA results by combined ELL and SWD status. However, Figure 3 compares 2007 NAEP fourth and eighth grade reading achievement scores for ELLs who either are or are not SWD. The scores lie well below NAEP's lower limits for the Basic achievement level of 208 in the fourth grade and 243 in the eighth grade, indicating that these students do not demonstrate an understanding of the overall or literal meaning of what they read.

Figure 3: English Language Learner Reading Results



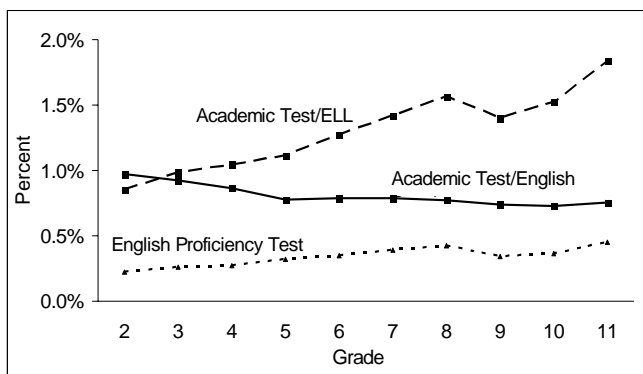
Both grades exhibit substantial gaps between SWD and non-SWD students. Given a standard deviation of about 50, the 35 point gap in the fourth grade and 27 point gap in the eighth grade, yield estimated effect sizes of .7 and .5, respectively. NAEP

scores are not on a common scale and do not support inferences about growth across grades. (Grigg and Donahue, 2007) Despite limitations, the NAEP achievement scores are consistent with California's results in Figure 2, further documenting a substantial gap in performance between ELLs who are SWD versus those who are not. Of course, any comparison of NAEP to California should consider that all SWDs and ELLs participate in California's testing program, whereas participation in NAEP varies by state.

Participation in Alternate Assessments

Figure 4 displays participation in alternate assessment for the CELDT and the CST in grades 2 through 11. Students who required an alternate assessment for the CST-ELA in 2007 took the CAPA. In grades two through eleven 3,994 ELLs participated in the CELDT alternate assessment and 13,787 took the ELA portion of the CAPA. One reason for the difference may be that the CST-ELA is an academic achievement test, and is more difficult than the CELDT. It is unlikely that the different testing windows (fall for the CELDT, spring for the CAPA) accounts for the discrepancy.

Figure 4: Estimated Participation in Alternate Assessment



The dotted line with triangles estimates participation of ELLs in alternate assessment on the CELDT. The percentages were calculated by dividing the total number of takers into the number who received an alternate assessment. The solid line with squares estimates participation of non-ELLs in alternate assessment on the CAPA. The percentages were calculated by dividing the number of English speakers (English Only and Fluent English Proficient students) who took the CST-ELA into the corresponding number who took the CAPA. The non-ELL line dips from the second to the fifth grade

and then displays a shallow downward trend. The participation of ELLs in the CAPA (dashed line with squares) rises more steeply than the CELDT line, but otherwise the two lines mirror each other. Both ELL lines display increasing participation rates, compared to a decline in the non-ELL line. Because the underlying numbers of students are large, the standard errors of percentages are small and differences greater than .1 percent are statistically significant ($p < .001$), notably after the fourth grade.

DISCUSSION

Comparison of overall participation rates in ELL and non-ELL populations in Figure 1 suggests that identification of SWDs lags somewhat for ELLs compared to non-ELLs. A more detailed examination reveals that the percentage of ELLs with disabilities rises steadily across grades. Below the fifth grade it appears that some ELLs with disabilities may be overlooked. After the fifth grade the percentages of ELLs with disabilities continues to rise, while the representation of non-ELLs declines. It is possible that cultural and communication barriers mask disabilities in younger ELLs. After teachers become better acquainted with the students, more accurate diagnosis of disabilities may become possible, resulting in higher rates of identification. In the upper grades it is also possible that continued lack of progress in gaining English proficiency by some students is misconstrued as evidence of cognitive impairment.

Given that many ELL/SWDs perform poorly on the CELDT, some may be inappropriately diagnosed as learning disabled, a label that covers a wide range of more or less well defined problems. Under the best of circumstances it is not easy to discriminate the possible deficits in a school's academic program from a student's learning disability. Differences in language and culture further complicate accurate diagnosis. An overabundance of care for the student, combined with an awareness of increased funding that is provided for SWDs, could result in a tendency to falsely identify students, who are not disabled but only need more appropriate or intensive language instruction. Despite good intentions, such misclassification could lower expectations and worsen the long-term academic outcomes for some ELLs. If so, this would be similar to the situation found in

Chicago's accountability program. (Roderick, Nagaoka, and Allensworth, 2007)

Figure 2 documents that ELL/SWDs score substantially lower than ELLs on the CELDT. The gap has an important consequence because the CELDT is one of the criteria used to decide whether to reclassify a student as FEP. Lower scores for ELLs who are correctly classified as disabled are not surprising. Such students not only face the difficulty of becoming proficient in English, they also require support for their disabilities. However, ELLs who are incorrectly classified as disabled may be placed in programs that offer lower expectations and inappropriate supports. The results suggest that classification of ELLs as disabled after the fifth grade deserves scrutiny. It would be helpful to study in more detail the characteristics of these students, their needs, and the services provided.

The findings for language proficiency testing reappear with academic achievement testing. The NAEP reading test results in Figure 3 document a substantial achievement gap between ELLs who are SWD or non-SWD. Considered separately, either ELL status or SWD status is an impediment to achievement. Poor language proficiency decreases access to curriculum and instruction. Learning or other cognitive disabilities pose a different, but no less potent, barrier. It should not be surprising that the combination is associated with even lower achievement. Where a true cognitive disability exists, students should receive whatever support or modified instruction and curriculum is appropriate. However, if an ELL is mistakenly identified as SWD, the danger exists that misdirected services may be ineffective, or worse.

ELL participation in alternate assessment increases across grades for the CELDT and the CST, with higher levels for the CST. One hypothesis relates to the difficulty of the tests. It may be that IEP team decisions for ELLs factor in a judgment of the difficulty of the test. The more difficult the test, the more likely is a decision in favor of alternate assessment. If true, this hypothesis could also help to explain the rise in participation across grades. The hypothesis is not supported for non-ELLs, for whom participation in the CAPA slightly declines across grades.

The alternate assessment participation rates for the CELDT and the CST double from grades two to eleven. The question arises whether the increase is related to a perverse incentive accompanying accountability for test scores. Students who take the easier alternate assessments do not take the more difficult regular tests and pose less of a threat to a school's accountability status. Testing for accountability where there are potentially harsh consequences for schools or students demonstrates the law of unintended consequences. NCLB punishment for failing to make adequate progress can involve decreased fiscal and program flexibility and eventual state takeover and reorganization of schools. One way of deflecting these consequences is to classify students as having a disability in the hope that they will receive additional assistance or will take an easier alternate assessment. Unfortunately, if the special education diagnosis is unsuitable, there is little assurance that these students will receive the instruction needed to improve their English proficiency skills.

REFERENCES

- Abedi, J. (2005). Issues and consequences for English language learners. In Herman, Joan L. and Haertel, Edward H. (Eds.). *Uses and Misuses of Data for Educational Accountability and Improvement. The 104th Yearbook of the National Society for the Study of Education. Part 2.* Malden, Massachusetts and Oxford, England. Blackwell Publishing.
- Abedi, J. (2007). English Language Proficiency Assessment and Accountability under NCLB Title III: An Overview. In Abedi, (Ed.) *English Language Proficiency Assessment in the Nation: Current Status and Future Practice.* University of California, Davis. Retrieved January 7, 2008 from http://education.ucdavis.edu/research/ELP_Assessment.html
- California Department of Education. (2007a). *California English Language Development Test (CELDT) Assistance Packet for School Districts and Schools.* Sacramento, California: California Department of Education. Retrieved January 7, 2008 from <http://www.cde.ca.gov/ta/tg/el>
- California Department of Education (2007b). *California Standards Tests (CST) Technical Report Spring 2006 Administration.* Sacramento, California: California Department of Education. Retrieved January 7, 2008 from <http://www.cde.ca.gov/ta/tg/sr/documents>
- California Department of Education. (2007c). *Dataquest.* Sacramento, California: California Department of Education. Retrieved January 7, 2008 from <http://dq.cde.ca.gov/dataquest/>
- California Department of Education (2007d). *CELDT Form E Technical Report.* Sacramento, California: California Department of Education. Retrieved January 7, 2008 from <http://www.cde.ca.gov/ta/tg/el/techreport.asp>
- The Commission on No Child Left Behind. (2007). *Beyond NCLB: Fulfilling the Promise to Our Nation's Children.* Washington, DC: The Aspen Institute. Retrieved February 10, 2007 from <http://www.aspeninstitute.org>
- Grigg, L. and Donahue P. (2007). *The Nation's Report Card: Reading 2007 (NCES 2007-496).* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C. Retrieved January 7, 2008 from <http://nces.ed.gov/nationsreportcard/pdf/main2007/2007496.pdf>
- Hoffman, L., and Sable, J. (2006). *Public Elementary and Secondary Students, Staff, Schools, and School Districts: School Year 2003-04 (NCES 2006-307).* U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved January 7, 2008 from <http://nces.ed.gov/pubs2006/2006307.pdf>
- Koretz, D., (2005). Alignment, high stakes, and the inflation of test scores. In Herman, Joan L. and Haertel, Edward H. (Eds.). *Uses and Misuses of Data for Educational Accountability and Improvement. The 104th Yearbook of the National Society for the Study of Education. Part 2.* Malden, Massachusetts and Oxford, England. Blackwell Publishing.
- Linn, R., (2005). Issues in the design of accountability systems. In Herman, Joan L. and Haertel, Edward H. (Eds.). *Uses and Misuses of Data for Educational Accountability and Improvement. The 104th Yearbook of the National Society for the Study of Education. Part 2.*

- Malden, Massachusetts and Oxford, England.
Blackwell Publishing.
- National Center for Education Statistics. (2008).
NAEP Data Explorer Results. Retrieved January
24, 2008
from <http://nces.ed.gov/nationsreportcard/nde/>
- Individuals with Disabilities Education Act (IDEA).
20 U.S.C. 1400 et seq. Public Law 108-446,
(1975/2004).
- No Child Left Behind Act of 2001 (NCLB). Public
Law No. 107-110., 115 Stat. 1425 (2002).
- Pullin, D. (2005). When one size does not fit all - The
special challenges of accountability testing for
students with disabilities. In Herman, Joan L. and
Haertel, Edward H. (Eds.). *Uses and Misuses of Data
for Educational Accountability and Improvement. The
104th Yearbook of the National Society for the Study of
Education. Part 2.* Malden, Massachusetts and
Oxford, England. Blackwell Publishing.
- Roderick, M, Nagaoka, J. and Allensworth, E. (2005).
Is the Glass Half Full or Mostly Empty? Ending
Social Promotion in Chicago. In Herman, Joan L.
and Haertel, Edward H. (Eds.). *Uses and Misuses of
Data for Educational Accountability and Improvement.
The 104th Yearbook of the National Society for the Study
of Education. Part 2.* Malden, Massachusetts and
Oxford, England. Blackwell Publishing.
- Tomalis, R. (2003). NCLB Policy Letter to New York
State. Office of Elementary and Secondary
Education. U.S. Department of Education.
Washington, D.C. Retrieved January 7, 2008
from <http://www.ed.gov/policy/elsec/guid/statelatters/asny.html>

Note

This research was supported by the JF Schouten School for User-System Interaction Research at Eindhoven University of Technology, Eindhoven, The Netherlands. The author thanks Steven Ralston for his language support.

Citation

Fetler, Mark (2008). Unexpected Testing Practices Affecting English Language Learners and Students with Disabilities Under No Child Left Behind. *Practical Assessment Research & Evaluation*, 13(6). Available online: <http://pareonline.net/getvn.asp?v=13&n=6>

Author

Mark Fetler
MarkFetler [at] gmail.com