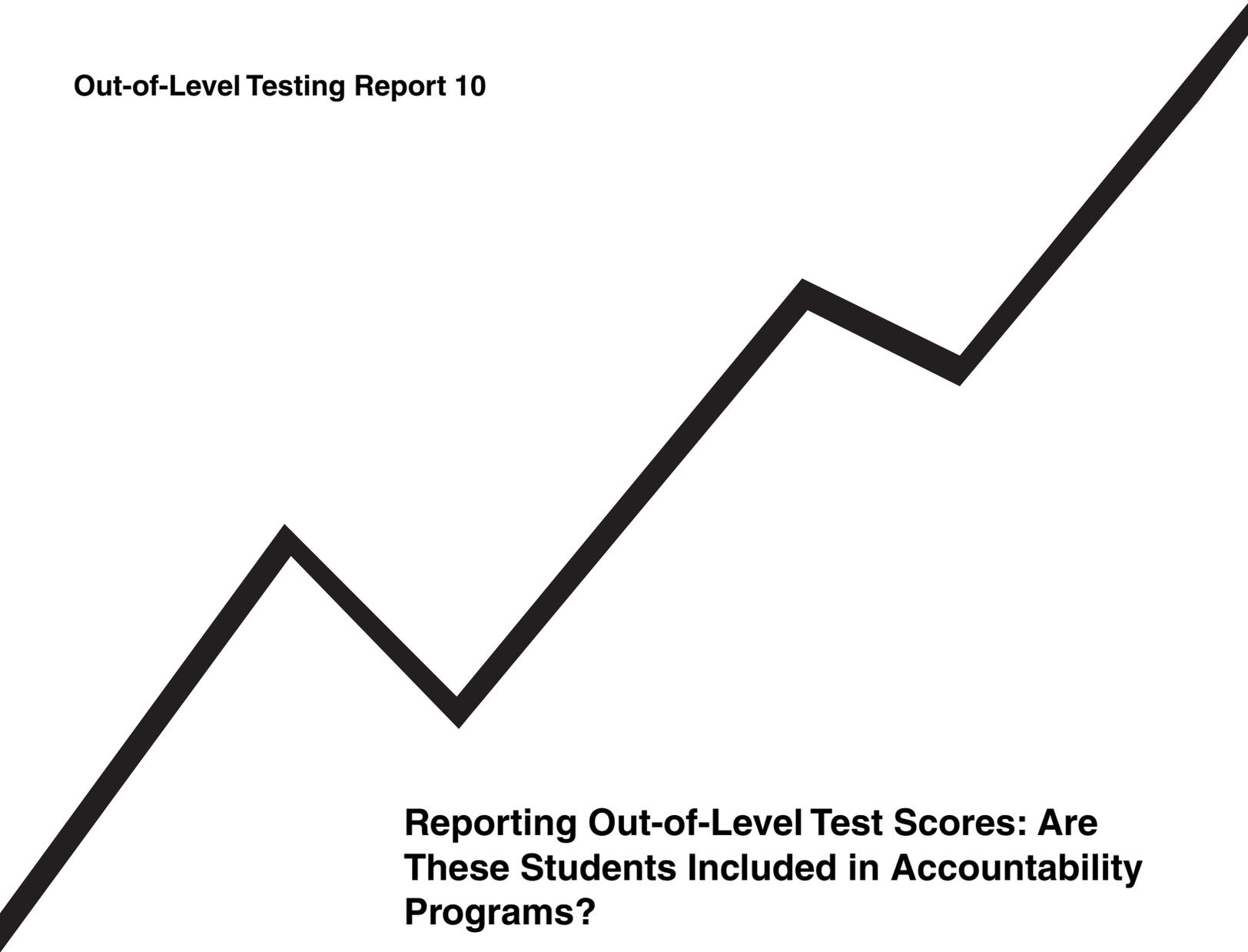


NATIONAL
CENTER ON
EDUCATIONAL
OUTCOMES

This document has been archived by
NCEO because some of the information
it contains may be out of date.

For more current information, please visit
NCEO's website at <http://nceo.info>



Reporting Out-of-Level Test Scores: Are These Students Included in Accountability Programs?



NATIONAL
CENTER ON
EDUCATIONAL
OUTCOMES

In collaboration with:

Council of Chief State School Officers (CCSSO)

National Association of State Directors of Special Education (NASDSE)

Out-of-Level Testing Report 10

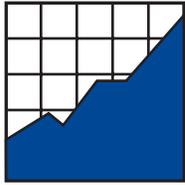
Reporting Out-of-Level Test Scores: Are These Students Included in Accountability Programs?

Jane Minnema • Martha Thurlow

October 2003

All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

Minnema, J., & Thurlow, M. (2003). *Reporting out-of-level test scores: Are these students included in accountability programs?* (Out-of-Level Testing Report 10). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.



**NATIONAL
CENTER ON
EDUCATIONAL
OUTCOMES**

The Out-of-Level Testing Project is supported by a grant (#H324D990058) from the Research to Practice Division, Office of Special Education Programs, U.S. Department of Education. Opinions expressed herein do not necessarily reflect those of the U.S. Department of Education or Offices within it.



NCEO Core Staff

Deb A. Albus	Ross Moen
Mike Anderson	Michael L. Moore
Ann T. Clapper	Rachel F. Quenemoen
Jane L. Krentz	Dorene L. Scott
Kristi K. Liu	Sandra J. Thompson
Jane E. Minnema	Martha L. Thurlow, Director

National Center on Educational Outcomes
University of Minnesota • 350 Elliott Hall
75 East River Road • Minneapolis, MN 55455
Phone 612/624-8561 • Fax 612/624-0879
<http://education.umn.edu/NCEO>

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.

This document is available in alternative formats upon request.

Executive Summary

With enactment of the No Child Left Behind (NCLB) Act of 2001, states are expected to ensure that all students are participating in a rigorous curriculum that is standards-based and on-grade level. States are also required to demonstrate adequate yearly progress, measured in part by large-scale assessment programs and made public through accountability data. In an attempt to create more inclusive large-scale assessment practices for students who have not been exposed to grade-level curriculum, some states have added out-of-level testing as a component of large-scale assessment programs. Out-of-level testing is the administration of a test at a level that is above or below the student's grade level in school. Typically, this means testing only students with disabilities below the grade in which their same-age peers are enrolled. However, because the intent of NCLB is to bring all students' achievement up to grade level standards, states are currently discouraged from testing any student below their grade of enrollment in school (Federal Register, July 5, 2002, pp. 45044-45).

In order to ensure that all students reap the benefits of participating in assessments, it is also necessary for states to enter every test score in their accountability system and report those results publicly. However, in a previous research study where we accessed states' large-scale assessment results from public reports, we encountered difficulties in locating out-of-level test data. In this study, we first accessed the data that were available in states' data reports for school years 1999-2000 and 2000-2001. Next we collected current analysis procedures used to prepare out-of-level test scores for public reporting. To provide the context within which these test results were reported, we also described the features of states' large-scale assessment programs for those states that tested students with disabilities out of level in statewide testing at the time of our data collection process.

Our findings paint a bleak picture of the status of publicly reported out-of-level test results. In our document review of states' data reports from school years 1999-2000 and 2000-2001, we were unable to locate any out-of-level test data that were clearly identified as students below the grade in which they were enrolled in school. Telephone interviews with state personnel indicated that some states are in the process of either partially or fully reporting out-of-level test scores. However, out-of-level test scores are still not clearly identified in states' data reports for those states that equated below-grade level test scores to on-grade level test scores. Two states reported that they did not report these test scores at all. Further analysis indicated wide variability in reporting practices across states when student, district, and state-level practices were compared. A qualitative analysis of the telephone interview data yielded four themes of results: (1) out-of-level test scores are not readily available within multiple types of state reports that contain large-scale assessment results, (2) few states have developed a process for reporting out-of-level testing results to the public, (3) out-of-level test scores that are reported publicly are not clearly identified as below grade level testing, and (4) states view reporting out-of-level test scores as a statistical problem.

We conclude the report by identifying three challenges in need of resolution before states can make informed decisions about out-of-level testing reporting practices: (1) there is a lack of consistency in states' out-of-level testing policies, which impedes the development of recommended guidelines for reporting test data; (2) quite often, state education agencies lack communication among different divisions, all of which are important to the successful implementation of large-scale assessment and accountability policy; and (3) there are various factors inherent to states' out-of-level testing policy that constrain reporting practices.

Out-of-Level Testing Background

Standards-based educational reform has taken hold across the nation. By the end of the 1990s, many policymakers and educators had championed the cause, and all but two states (Iowa, Nebraska) had developed and implemented large-scale assessment and accountability programs statewide. The expansion of statewide testing and accountability programs occurred in part to meet the legal requirements of the 1994 Elementary and Secondary Education Act (ESEA) and the 1997 reauthorization of the Individuals with Disabilities Education Act (IDEA 97). Both laws required that *all* students participate in states' testing programs, with the Title I legislation extending the mandate to include accounting for *all* students. With enactment of the No Child Left Behind (NCLB) Act of 2001, which is the most recent re-authorization of ESEA, states are expected to ensure that *all* students are participating in a rigorous curriculum that is standards-based and on-grade level. States are also required to demonstrate adequate yearly progress, measured by large-scale assessment programs and made public through accountability data.

Accordingly, today more than ever, the public wants students and schools to demonstrate improved educational results. As public scrutiny has increased, states have begun to realize that all students, especially those with disabilities, were being assessed in testing not used for accountability purposes. In an attempt to create more inclusive large-scale assessment practices for students who have not been exposed to grade-level curriculum, states added out-of-level testing as a component of large-scale assessment programs. Out-of-level testing is the administration of a test at a level that is above or below the student's grade level in school. Typically, this means testing only students with disabilities below the grade in which their same-age peers are enrolled. Once intended to measure program effectiveness of Title I interventions in the 1970s, the current use is one of measuring students' academic progress toward attaining states' content standards. In 2001-2002, 14 states (Arizona, California, Connecticut, Delaware, Hawaii, Iowa, Louisiana, Mississippi, Oregon, South Carolina, Texas, Utah, Vermont, West Virginia) tested students out of level in their large-scale assessments.

The use of out-of-level testing, or the administration of a test at a level lower than a student's age or grade level in school, has expanded within a contentious and politicized environment (Thurlow & Minnema, 2001). In fact, many states prefer not to use the term out-of-level testing, opting instead for terms that may invoke less adverse reactions. For instance, some states refer to below grade level testing as off-level testing, alternate assessment, alternative assessment, or challenge down testing. Because the current federal agenda for several years has been neither receptive nor supportive of the use of out-of-level tests for large-scale assessments, we acknowledge that the term "out-of-level testing" is not the preferred term for all of the states from which we collected data. It is also important to acknowledge that data were collected for this project prior to the enactment of No Child Left Behind in late 2001. That law clearly addresses the need for grade-level testing for all students' progress toward achieving grade-level

content standards. The law also states, “The U.S. Department of Education considers out-of-level testing as not an acceptable means for a state to meet its assessment requirements under NCLB although such tests might be appropriate for other purposes.” (Federal Register, July 5, 2002, pp. 45044-45). Given this, there may be instances where states’ policies have changed in response to federal regulations. These changes will not be evident in the results of the research reported here.

States justify testing students out of level by claiming that more students with disabilities participate in statewide testing when tested at their instructional level. In order to ensure that all students reap the benefits of participating, it is also necessary for states to enter every test score in an accountability index and report those results publicly. To date, research has only described the perceptions of state level personnel about the reporting of out-of-level test results (Minnema, Thurlow, & Scott, 2001). That study indicated that data managers found it difficult to include out-of-level test scores in statewide assessment reporting. When we began to look for data on the prevalence of out-of-level testing in statewide assessments by reviewing states’ public data reports, we found no disaggregated out-of-level test results clearly reported. This finding pushed us to request out-of-level test data directly from state educational agencies (Thurlow, Minnema, Bielinski, & Guven, 2003). In the end, only a few states supplied data.

Because of the difficulty we encountered in obtaining data for the 2000-2001 school year, we undertook a descriptive study of the reporting of out-of-level test data. This research had two purposes: (1) to describe what data were available in states’ data reports for school years 1999-2000 and 2000-2001, and (2) to describe current analysis procedures used to prepare out-of-level test scores for public reporting.

Out-of-Level Testing Practices

In order to understand how states report out-of-level test results, it is helpful to take a step back and consider the practices used when out-of-level tests are administered. Just as each state has created different statewide assessment and accountability programs, the practices used to implement out-of-level testing differ across states also. A recent comprehensive review of out-of-level testing policies, which was updated in 2002, yielded few similarities across states (Thurlow & Minnema, 2001). In fact, just one point of commonality among these out-of-level testing policies emerged. All 14 states that allowed out-of-level testing did so for students with disabilities. A few states test other subgroups of students below their grade of enrollment in addition to students with disabilities. There are few exceptions to this finding. Two states tested students with disabilities and students with 504 accommodation plans out of level (Utah, West Virginia), and

one state tested any student who met the state out-of-level testing criteria (Vermont). Taken all together, it is important to note that the majority of statewide tests that are administered out of level are for students who receive special education services. (See Thurlow and Minnema, 2001, for a more in-depth discussion of the contextual issues concerning out-of-level testing.)

Table 1 presents some of the features of states' large-scale assessment programs that are relevant for our discussion of reporting out-of-level scores. Note that this table incorporates the changes in states' out-of-level testing policies from 1999 through 2002. Throughout the school year 1999-2000, 12 states (Alaska, Arizona, California, Connecticut, Delaware, Iowa, Louisiana, North Dakota, South Carolina, Utah, Vermont, West Virginia) tested students out of level in large-scale assessment programs. Since then, two states (Alaska, North Dakota) have discontinued an out-of-level testing policy while four states (Hawaii, Mississippi, Oregon, Texas) have initiated some version of testing students below grade level in their statewide assessment programs.

Each of these states administered a different type of testing instrument. Some states used a norm-referenced test (Iowa, North Dakota, West Virginia), some used a criterion-referenced test (Alaska, Connecticut, Mississippi, Oregon, South Carolina, Texas, Utah, Vermont), and others used a combination of a norm-referenced and criterion-referenced test (California, Delaware, Hawaii). In addition, one state (Arizona) tested out of level with both a CRT and an NRT while one other state (Louisiana) tested with an NRT for students who did not pass the CRT.

Few states treated out-of-level testing similarly in their state level assessment policies. Four states (Alaska, Arizona, South Carolina, West Virginia) considered out-of-level tests as modifications to a standard test presentation while three states (Delaware, Hawaii, North Dakota) treated out-of-level testing as an accommodated test. California labeled out-of-level testing a non-standard accommodation. The remaining states used a variety of labels for testing students below the grade level in which students are enrolled in school, including alternate assessment option #1, alternate assessment, benchmark challenge test, instructional level test, alternative test, and adapted (out-of-level) alternate assessment.

In terms of accountability programs, few states incorporated out-of-level test scores into their longitudinal measurement of groups of students' progress toward achieving state content standards. For those states (Iowa, North Dakota) that do, both administered an NRT where out-of-level test scores could be equated to on-grade level test scores for use in an accountability index. One other state (Vermont) developed transformation rules that convert below-grade level scores to on-grade level scores for accountability purposes. There are no states that used out-of-level test scores for making high stakes decisions for either students or school systems.

Table 1. Out-of-Level Testing Features by State

State	Type of Instrument	OOLT Classification	Equate to In-Level Scores	Accountability System(s)
<i>Alaska*</i>	<i>CRT</i>	<i>Modification</i>	<i>No</i>	<i>Student accountability with voluntary system accountability</i>
Arizona	CRT NRT	Modification	Not determined	Student accountability by 2002
California	NRT/CRT	Standard (1 level below) & Non-standard (2 or more levels below) Accommodation	No	Student and system accountability
Connecticut	CRT	Alternate Assessment Option #1	No	Student and system accountability
Delaware	CRT/NRT	Accommodation	No	Student and system accountability
Hawaii**	CRT/NRT	Accommodation	No	School and system accountability
Iowa	NRT	Alternate assessment	Yes (Could equate)	None
Louisiana	NRT (In lieu of CRT)	Alternate assessment	Disaggregated	Student and system accountability
Mississippi**	CRT	Instructional level test	No	Student, school, and system accountability
North Dakota*	<i>NRT</i>	<i>Accommodation</i>	<i>Aggregated</i>	<i>School accountability</i>
Oregon**	CRT	Challenge down	No	Student, school, and system accountability
South Carolina	CRT	Modification	Disaggregated	System accountability
Texas**	CRT	Alternative test	Disaggregated	Student, school, and system accountability
Utah	CRT	Alternate Assessment	Disaggregated	System accountability (At district level by submitting accreditation report to Northwest Association of Schools & Colleges) Student accountability by 2005
Vermont	CRT	Adapted (Out-of-Level) alternate assessment	Equated scores entered in accountability index	Student and system accountability
West Virginia	NRT	Modification	Aggregated with all non-standard SAT-9 scores	Student and system accountability

* States eliminated out-of-level testing, 2001.

** States initiated out-of-level testing, 2001

Method

Our purposive sample included all data reports from states that indicated that out-of-level tests were administered to students with disabilities as a component of statewide testing during the 1999-2000 testing cycle (Bielinski, Thurlow, Callender, & Bolt, 2001). We used two data collection strategies for gathering information from two sources of data. First, to understand what data were publicly reported in states' data reports, NCEO researchers conducted systematic reviews of data reports that were downloaded on the World Wide Web or accessed directly from state education agencies (SEAs). Publicly reported data were collected for both the entire group of students with disabilities and also for the subgroup of students with disabilities who were tested out of level. The document reviews were conducted on the results of states' large-scale assessments for both school years 1999-2000 and 2000-2001. (See Bielinski, Thurlow, Callender, and Bolt, 2002 and Thurlow, Wiley, and Bielinski, 2003 for a more thorough discussion of these assessment results.)

A second data collection activity involved direct contact with SEA personnel to learn about how out-of-level test results are prepared and posted for public reporting. As a first step, we reviewed Thurlow and Minnema (2001) to document how states described their reporting practices for school year 2000-2001. To update this information, we conducted telephone interviews with state level personnel (n = 16) who were recommended to us as familiar with reporting large-scale assessment results. Our telephone interview protocol questions included:

1. Please describe the process used by your school districts to submit out-of-level test scores to your state educational agency.
2. How are out-of-level test scores reported at the local level? Are these data made public?
3. How are out-of-level test scores reported at the state level? Are these data made public?
4. Please describe the analysis procedures used for public reporting of out-of-level test scores.

Probe: Are out-of-level test scores aggregated? If so, with what other scores? At what grade level?

Probe: Are out-of-level test scores disaggregated? If so, what categories are used to disaggregate these data (e.g., disability category, grade level tested, assigned grade level)?

Each telephone interview was tape recorded. To analyze the narrative data, we listened to the tape recordings to glean relevant information that would address each interview question for each state that allowed out-of-level testing during the school year 2000-2001. To finalize our

results, we conducted a semi-structured content analysis to identify general patterns in the interview data.

Results

Our results are presented in two sections, with the analysis of data reports first, followed by the narrative results of our telephone interviews with SEA personnel. In the final portions of this report, we interpret and discuss our findings by treating the two data sets as a composite whole set of results.

Analysis of States' Data Reports for Out-of-Level Test Results

We present the findings of our review of states' reports on large-scale assessment results in Table 2 (for the school year 1999-2000) and in Table 3 (for school year 2000-2001). An indication of whether data were located for the participation and the performance of students with disabilities in statewide testing is also presented. This provides a context for understanding what we found on out-of-level testing reporting. If our document reviews did not yield any out-of-level test results, we indicate that finding with "unable to locate."

For the school year 1999-2000, seven of the states using out-of-level testing (Arizona, Connecticut, Iowa, Louisiana, North Dakota, Utah, West Virginia) disaggregated data by conducting analyses on student subgroups for both the performance and participation of students with disabilities in their states' large-scale assessment. Three states (Delaware, South Carolina, Vermont) disaggregated performance results only; one state (California) included no disaggregated data for students with disabilities in its state data report. In terms of reported results for the participation and performance of students with disabilities who were tested out of level in statewide testing, no state included those test data when reporting to the public. Some states may have equated below-grade level test scores to on-level scores, and then reported on the student's grade of enrollment. However, since these test results were not labeled "out-of-level," we were unable to locate them. Please note that Alaska was not included in this table because out-of-level testing was allowed for one school year, and then only for English language learners who attended a 4th grade language immersion program. No other students were to be tested out of level in school year 1999-2000. After that school year, no students were to be tested below grade level in Alaska's statewide assessment program.

Of the 14 states that tested students with disabilities out of level in statewide testing programs during the school year 2001-2002, 3 states (Arizona, Hawaii, Vermont) reported no disaggregated data for students with disabilities in either the regular assessment or the state assessment

Table 2. Statewide Test Data for Students with Disabilities for 1999-2000

State	Report on ALL Students with Disabilities	Report on Out-of-Level Tests
Arizona	Performance and participation data disaggregated	Unable to locate
California	No disaggregated data	Unable to locate
Connecticut	Performance and participation data disaggregated	Unable to locate
Delaware	Performance data disaggregated	Unable to locate
Hawaii*		
Iowa	Performance and participation data disaggregated	Unable to locate
Louisiana	Performance and participation data disaggregated	Unable to locate
Mississippi*		
North Dakota	Performance and participation data disaggregated	Unable to locate
Oregon*		
South Carolina	Disaggregated performance data	Unable to locate
Texas*		
Utah	Performance and participation data disaggregated	Unable to locate
Vermont	Disaggregated performance data	Unable to locate
West Virginia	Performance and participation data disaggregated	Unable to locate
*Not testing out of level 1999-2000.		

administered below grade level. We found performance and participation data disaggregated for students with disabilities for eight states (California, Connecticut, Delaware, Iowa, Mississippi, Oregon, Texas, Utah). Three states (Louisiana, South Carolina, West Virginia) did not disaggregate test results for all components of their statewide testing programs. Of the eight states with disaggregated data, only Connecticut made out-of-level test results public by reporting the number of students with disabilities who participated in out-of-level testing. No state's data reports contained clearly labeled out-of-level test performance results for students with disabilities.

Analysis of Telephone Interview for States' Reporting Practices

Table 4 summarizes the results from two data collection activities that involved telephone contacts with SEA personnel regarding two testing cycles. For school year 1999-2000, Thurlow and Minnema (2001) found wide variability in how states managed and reported the results

Table 3. Statewide Test Data for Students with Disabilities from 2001-2002

State	Reported for ALL Students with Disabilities	Reported for Out-of-Level Tests
Arizona	No disaggregated data	Unable to locate
California	Performance and participation data disaggregated	Unable to locate
Connecticut	Performance and participation data disaggregated	Disaggregated participation data
Delaware	Performance and participation data disaggregated	Unable to locate
Hawaii	No disaggregated data	Unable to locate
Iowa	Performance and participation data disaggregated	Unable to locate
Louisiana	Performance and participation data disaggregated (but no disaggregated data for the Developmental Reading Assessment)	Unable to locate
Mississippi	Performance and participation data disaggregated	Unable to locate
Oregon	Performance and participation data disaggregated	Unable to locate
South Carolina	Disaggregated performance data; disaggregated participation data for the PACT	Unable to locate
Texas	Performance and participation data disaggregated	Unable to locate
Utah	Performance and participation data disaggregated	Unable to locate
Vermont	No disaggregated data	Unable to locate
West Virginia	Performance and participation data disaggregated (but no disaggregated data for the Writing Assessment)	Unable to locate

from out-of-level tests. No two states used the same procedures for including out-of-level test scores in accountability and then analyzing those data for public reporting purposes. Four states (Delaware, Louisiana, South Carolina, Utah) indicated that out-of-level test data were disaggregated (i.e., analyzed separately for students with disabilities). In doing so however, one state (Delaware) disaggregated the test data without reporting norm-referenced test scores at the state level. Four other states (Iowa, North Dakota, Vermont, and West Virginia) reported aggregating out-of-level test results for reporting to the public. However, only two states (Iowa, North Dakota) did so in a like manner by equating out-of-level test scores to on-grade level test scores for norm-referenced tests. Vermont had developed a system of transforming out-of-level test scores to on-grade level test scores for their criterion-referenced state test while West Virginia combined all types of nonstandard test results in one aggregated score. One state (California), which treats an out-of-level test as a nonstandard test presentation if the test level is more than one grade level below, reports no nonstandard test results at the state level. An-

Table 4. Out-of-Level Testing State Level Reporting Practices According to SEAs

State	For School Year 1999-2000 *	For School Year 2001-2002 **
Arizona	In development	Not reported
California	Standard scores aggregated with grade of enrollment with nonstandard scores not reported at state level	Standard scores aggregated with grade of enrollment with nonstandard scores not reported at state level
Connecticut	Only participation reported	Participation reported for grade level of test
Delaware	Disaggregated without reporting NRT scores	Aggregated at lowest proficiency for grade of enrollment
Hawaii	Not testing out of level	Reporting procedures in development
Iowa	Aggregated	Aggregated with grade of enrollment
Louisiana	Disaggregated	Aggregated at grade of enrollment
Mississippi	Not testing out of level	Aggregated at lowest proficiency level for grade of enrollment (Writing Test only)
North Dakota	Aggregated at grade of enrollment	No longer testing out-of-level
Oregon	Not testing out of level	Reporting procedures in development
South Carolina	Disaggregated	SEA not responsible for report of modified tests in district or state reports
Texas	Not testing out of level	Disaggregated performance by grade, test level, and demographic groups
Utah	Disaggregated	Participation reported for grade level of test
Vermont	Equated scores entered in accountability index	Not reported

* Source: Thurlow & Minnema, 2001.

** Source: Summer, 2001 data collection; Updated Fall, 2002.

other state (Arizona) indicated that reporting procedures for including out-of-level test results in data reports made public were in the process of being developed. Please note that four states (Hawaii, Mississippi, Oregon, Texas) had not fully initiated an out-of-level testing policy for school year 1999-2000.

The results of the second phase of data collection indicated that reporting practices had not changed over the school years in four states (California, Connecticut, Iowa, West Virginia). In

comparing the reporting practices between the two school years, four states changed the way in which out-of-level test scores were reported (Delaware, Louisiana, South Carolina, Utah). Each state had more reporting-specific information at the second point in data collection. Delaware reported out-of-level test scores aggregated at the lowest proficiency level for the grade of enrollment regardless of test performance on the level at which tested. Louisiana equated out-of-level test scores to on-grade level test to report in aggregate on the grade of enrollment. South Carolina SEA does not report out-of-level tests scores because an independent agency develops the state data report for state accountability purposes. The fourth state that had changes in reporting practices, Utah, reported participation test data on the grade at which students were tested.

Two states did not report out-of-level test scores (Arizona, Vermont). In the case of Vermont, test-specific transformation rules have been developed to equate out-of-level test scores to on-grade level test scores. However, the validation of the rules has not yet been completed so that statewide test scores are not yet reported. States that were new to testing students below grade level by school year 2001-2001 were in the process of developing reporting practices. Three of these states (Mississippi, Oregon, Texas) were able to describe the point at which reporting procedures had been developed. Mississippi had reported Writing Test scores in aggregate at the lower proficiency with plans to do so for all content areas tested out of level in the future. Oregon did not report challenge down test scores for students with disabilities for the school year 2001-2002, but planned to do so in the future. Texas disaggregated performance test data for below-grade level testing for the alternative state test according to grade, test level, and demographic groups. Plans were in place to report progress toward proficiency levels in the future for those students with disabilities tested below-grade level. The final state new to out-of-level testing, Hawaii, did not have reporting information at this point in time. One state (North Dakota) had ceased testing students with disabilities out of level by the school year 2001-2002.

Processing Test Scores

Responses to telephone interviews indicated that, with the exception of one state (Hawaii), all states received local level out-of-level test results in a similar manner. SEAs received test results via state contracts with testing vendors and designated dates for test administrations, but school districts were responsible for coordinating the administration of state tests and returning the answer sheets to the test company. Local educational agencies (LEAs), most typically at the district level, requested the number of out-of-level tests by test level needed per school for each testing cycle from the state's test contractor. Once the tests are administered, an educator who serves as a district test coordinator collects and packages the tests to return to the test company. Test companies scan test bubble sheets yielding test scores. Raw test data files are then submitted

to the SEA for data analysis. Test results are distributed to LEAs as well as published in states' data reports for public examination.

Because of the unique organization of the Hawaii school district, scores are only transferred between the test contractor and the state assessment director who functions as the district test coordinator. The state in its entirety is one school district so that one administrator serves as both state assessment director and district test coordinator. State test data made public at the local and state level are published in one district/state report.

Reporting Out-of-Level Test Scores at Local and State Levels

Table 5 displays the wide variability in the procedures that states use to report out-of-level test results, as reported in our interviews with states. Since some states do not necessarily provide similar information at the student and district level, we made these distinctions for local level reporting.

Only two states (Iowa, Louisiana) have adopted similar reporting procedures at the district and state level, probably because both states administer the same standardized instrument out of level. Both of these states report out-of-level test results in aggregate on the grade of enrollment. The states differ on reporting at the student level. Louisiana distributes an individual student report while Iowa indicated that it was considering doing so in the future. Other states use different approaches.

Three states (Arizona, Vermont, West Virginia) did not fully report out-of-level test scores. Arizona reported student performance on the grade of enrollment to students and parents, while not reporting these test results at the district or state level. Vermont also only reported out-of-level test results at the student level, but distributed school reports rather than individual student performance. West Virginia used yet another set of reporting procedures whereby out-of-level test scores were reported at the state level, but not at the local level.

Only two states (Texas, Utah) used similar reporting procedures across student, district, and state levels. Texas disaggregates all below grade level test results. At the student level, test results are disaggregated by content area while at the district and state levels data are disaggregated by both grade of enrollment and level at which tested. Other states (Connecticut, Delaware, South Carolina) vary both across and within states in terms of how out-of-level test scores are reported. Connecticut only reports participation data at the state level while reporting performance results by test level at the student level. The SEA does distribute a separate document to all schools in Connecticut that list students tested below grade level by grade of enrollment, level at which tested, and test score. Delaware and Oregon report all out-of-level test scores at the lowest level of proficiency on grade-level standards for both district and state reporting. Student performance

Table 5. Out-of-Level Test Reporting Practices by Student, District, and State

State	Student	District	State
Arizona	Reported separately on grade level.	Not reported	Not reported
California	Standard test scores reported as on-grade level scores. Nonstandard raw scores only.	Only standard scores equated, aggregated, and reported on grade of enrollment. Nonstandard not reported	Only standard scores equated, aggregated, and reported on grade of enrollment. Nonstandard not reported
Connecticut	Performance reported by test level	List of students by grade of enrollment, by test level, and score	Participation reported by test level
Delaware	Performance reported on grade level tested to parents	Reported at lowest proficiency level at grade of enrollment	Reported at lowest proficiency level at grade of enrollment
Hawaii*	Procedures in development	Procedures in development	Procedures in development
Iowa	Considering individual student reports	Performance equated, aggregated, and reported on grade of enrollment	Performance equated, aggregated, and reported on grade of enrollment
Louisiana	Individual student report	Performance equated, aggregated, and reported on grade of enrollment	Performance equated, aggregated, and reported on grade of enrollment
Mississippi*	Procedures in development	Participation report by test levels for writing	Procedures in development
Oregon	Performance reported on benchmark level to students and teachers	Performance aggregated at lowest proficiency level on benchmark level	Performance aggregated at lowest proficiency level on benchmark level
South Carolina	Modified test score reported to parents	Performance aggregated for state report card with %age tested out of level.	Not reported in state data report made public
Texas*	Disaggregated student performance by content area	Disaggregated performance by grade, test level, and demographic groups	Disaggregated performance by grade, test level, and demographic groups
Utah	Participation reported by grade level tested.	Participation reported by grade level tested	Participation reported by grade level tested
Vermont	School reports to parents	Not reported	Not reported
West Virginia	Not reported	Not reported	Aggregated with all nonstandard scores

* First year of testing below grade level.

according to the test level is reported at the student level in these two states. Differing further, out-of-level test data are reported by two organizations in South Carolina, each of which uses different procedures. At the student level, an individual score report is distributed to teachers and families indicating a modified assessment that does not necessarily represent achievement toward grade-level standards. The SEA also prepares a state data report for the public in which out-of-level test scores are not reported. For accountability purposes in South Carolina, an Education Oversight Committee prepares a state report card where out-of-level test scores are aggregated by level of proficiency regardless of the grade level at which students are tested.

At the time of our data collection activity, three states (Hawaii, Mississippi, Texas) were implementing an out-of-level testing policy for the first time. Table 5 contains information from each states' first testing cycle. We explain probable next steps in reporting out-of-level test results in each of these states for as much as the interviewees were able to project.

Hawaii, a state that is comprised of one school district, is in the process of determining how to report out-of-level test scores. At the time of our data collection, future plans were to report out-of-level test scores in a matrix of all scores that is distributed to each school. In order to report at the state level, test scores are tabulated across all schools. It was thought that an out-of-level test score would be reported at the lowest proficiency level, indicating that grade-level standards had not been met.

A second state, Mississippi, was also in the first year of testing out of level. In fact, at that point in time, the SEA had not yet received results from the Mississippi Curriculum Test because the process to set grade-level, content standards had not been completed. The SEA did report on the writing test, which was administered earlier, by submitting summary reports to school districts. These reports contained aggregated participation data on the writing test according to the grade levels at which students were tested. Specific decisions about the configuration of aggregated and disaggregated test data for other content areas are in process. The SEA plans to make these data public as well as providing individual score reports for teachers and parents.

Texas was also between the first and second testing cycles at the time of our data collection. Table 5 contains reporting information for the first year of testing. School, district, and state level alternative test data were reported, with the number participating disaggregated by grade level tested for each grade of enrollment in school. In addition, by using the results from this testing cycle as baseline data, a percentage of students who meet academic growth across two years of testing as projected by their assessment, referral, and dismissal committees (ARDs) will be reported. However, the interviewee from the Texas state educational agency indicated that the details for these practices were not fully determined yet.

Public Reporting

Reporting large-scale assessment results as a measure of academic progress toward grade level standards is in various stages of development across states that test students out of level. SEA personnel who participated in our telephone interviews described procedures that varied widely from state to state. In fact, the most common reporting feature across states was that no state is currently making out-of-level test scores public information in a clearly identifiable manner. On the other hand, one of the most promising findings from our telephone interviews is the finding that out-of-level data are being reported. The issue then becomes not necessarily one of whether out-of-level test scores are reported, but rather how these data are reported. There are two specific issues that emerge from a deeper understanding of publicly reported out-of-level test results. The first issue is one of what specific statistics are reported. The second issue relates to the procedures by which test data are included in states' data reports.

First, in terms of the type of data reported, three states (Connecticut, South Carolina, Utah) provide participation data for out-of-level tests in such a way that the public knows how many and what percentage of students are tested below their grade of enrollment. While participation rates are important indicators, full disclosure of performance data is just as important. Participation rates are an admirable first step to increased reporting, but not the end point in accounting for achievement over time.

Our findings also indicated that no state disaggregates out-of-level test results by disability category. This type of analysis could be informative for policymakers, teachers, and parents. Disproportionate participation of specific disability subgroups in out-of-level testing would have important ramifications for states that are striving to include all students in states' accountability programs.

A second public reporting issue, how states manipulate data for accountability purposes, points to the complexities involved in using out-of-level test results for accountability purposes. To find these results in states' data reports, it is necessary to describe reporting procedures specifically because the out-of-level test data are masked by the manner in which the results are reported. For instance, Iowa uses a norm-referenced test so that the out-of-level test scores can be equated to grade-level test scores for reporting results on students' grade of enrollment. Delaware reports all out-of-level test scores on the lowest proficiency level, indicating that students striving to achieve lower grade level standards are "below basic" in achieving grade level standards. Another state, West Virginia, treats out-of-level tests as nonstandard test presentations, so that out-of-level test scores are reported in aggregate with other nonstandard test scores. Technically, in each of these cases, out-of-level test scores are reported in states' data reports. These results may be posted on SEAs' Web sites. However, it is not possible to know

where the out-of-level test scores are incorporated into the statistical analyses unless the details of the procedures are specified.

There are also some states that either partially report out-of-level test scores or do not report these scores at all. Since California treats out-of-level tests that are administered one level below grade level as standard test administrations and more than one level below grade level as nonstandard test administrations, only standard test scores are reported. The remaining out-of-level tests in California are not reported. Three states (California, Vermont, West Virginia) indicated during our telephone interviews that their out-of-level test data are not made public. There may be additional states that fall into this category, but were unwilling to indicate so when interviewed. Two more states (Hawaii, Mississippi) were in the process of finalizing reporting practices for out-of-level testing at the time of our data collection activity.

Purpose of Reported Score

Across states, out-of-level tests do not necessarily serve the same purpose. For most states, test results are intended to be a measure of academic progress toward content standards that are developed for a lower grade level than the grade in which a student is enrolled in school. However, in two states (Texas, Delaware) out-of-level test scores are not necessarily indicators for specific grade levels of content standards. In Texas, ARD committees determine a projected amount of progress toward content standards that students achieve throughout a school year. Scores from below grade level testing serve as a measure of growth to determine whether students meet the projected achievement, and are then reported as such. In contrast, out-of-level test data in Delaware do not represent specific academic achievement. The rationale for assigning the lowest proficiency level for all out-of-level test scores regardless of the statewide test score is to indicate that students have not reached proficiency on grade-level standards. Teacher and families receive more specific information as performance at the test level is shared at the student level. Schools receive some credit for a student who is achieving below grade level by reporting at the lowest proficiency, but remain responsible for students who are not achieving at the grade in which they are enrolled.

Future Procedures

In our interviews, two states indicated projected changes in out-of-level testing policy for school year 2002-2003. Iowa indicated awareness of the importance of determining the number of students tested statewide below grade level as well as the grade levels at which students are tested. The SEA planned to put procedures in place in order to report on the numbers of students tested

below grade level. It was also indicated that Arizona is moving toward reporting out-of-level test participation rates at the state and local level.

Qualifying Statements

Most states provided various rationales for the non-report of out-of-level test data. One state (Vermont) indicated that the number of students tested out of level statewide was so few that omitting these test scores from test data aggregated at the state level had a negligible effect on the resulting numbers. Disaggregating out-of-level test performance and participation had little meaning in Vermont, again because of the limited use of below grade level testing. Other states, Connecticut and Utah in particular, indicated that aggregating out-of-level test scores was not feasible since their statewide test was criterion referenced. Without a common scoring scale, out-of-level test scores cannot be meaningfully equated to on-grade level test scores. Finally, three states (Hawaii, Mississippi, Louisiana) specifically mentioned that reporting out-of-level test scores in a public format was not possible if the grouping of students was less than 10 students where confidentiality could potentially be violated. In Louisiana, for instance, school reports are distributed with out-of-level test data only for those schools that test 10 or more students out of level.

Discussion of Issues

Every effort was made to incorporate the most current information about reporting practices for those states that allow out-of-level testing in their large-scale assessment and accountability programs. Wherever possible, we used personnel from state educational agencies whose role is directly related to assessment and accountability programs as our source for data. Even so, it is possible that the interviewees may not have had complete information on reporting practices in their state. It is also possible that our information may not be fully updated if policy changes occurred as this report was being prepared. With that understanding, we present four central issues that evolved from our review of out-of-level testing policies.

Issue #1—Out-of-level test scores are not readily available within multiple types of state reports that contain large-scale assessment results.

Generally speaking, the results of statewide tests that were administered below the grade in which students were enrolled in school were not readily accessible in either states' data reports or states' Web sites. If states are going to use out-of-level testing, it is imperative that they

develop procedures to clearly report out-of-level test results to the public. Reporting in aggregated as well as disaggregated form is equally important. This means that scores for students with disabilities are included in “all students” results as well as in “students with disabilities tested below enrollment grade level.” In order for students with disabilities to reap the benefits of school improvement planning, it is necessary to accurately count and meaningfully consider their test performance. It is only by disaggregating test participation and performance that states can monitor how many students are tested below grade level as well as how well these students are challenged by state tests.

There are two states that have made an inroad toward clear reporting of out-of-level test data. In South Carolina, for instance, the results of out-of-level testing are reported in an individual student report as a modified assessment that describes the progress made toward standards at the grade of enrollment (a grade that is different from the one in which the student is tested). As another example, Connecticut provides a separate report to districts with detailed information about out-of-level testing in their schools.

Issue #2—Few states have developed a process for reporting out-of-level testing results to the public.

The purpose of publicly reporting large-scale assessment results is to provide an accounting of schools’, districts’, and states’ progress toward achieving grade-level content standards. However, when students are tested below grade level, including these test data in accountability indexes is very complex. Questions arise such as whether out-of-level test results should be reported at the grade level of testing or enrollment. If reported at the testing grade level, what does that say about achieving content standards at the grade of enrollment? Or, if the test data are reported at the grade of enrollment, what information does that provide about students’ proficiency on a set of content standards that are below-grade level?

In response to these concerns some states, such as Delaware and Texas, have developed unique procedures for including below-grade level test scores in accounting for academic progress. By doing so, instructional questions arise for students in those states that do and do not report out-of-level test results. What happens to the academic progress of students who are tested below grade level over consecutive school years? How does out-of-level testing affect the learning expectations set by teachers, parents, and the students themselves? What happens to the graduation and dropout rates when students with disabilities are tested out of level at young ages? These are important issues that need to be resolved in order for states to be in compliance with current legal mandates.

Issue #3—Out-of-level test scores that are reported publicly are not clearly identified as below grade level testing.

For those states that use a norm-referenced instrument for statewide testing, test companies have developed normative data to equate below-grade level test scores to on-grade level test scores. In these cases, out-of-level test scores are reported on the grade at which a student is enrolled in school. However, since test scores are transformed to grade level scores, it is not possible to know how many students were tested at which grade level. It is also not possible to determine what test performance was according to the grade at which the students were tested. Since these data are not disaggregated by any variable, states' data reports do not convey student characteristics for those students who are tested out of level. We believe these data to be especially critical to describing the results of statewide testing because students with disabilities are typically those students who are tested below grade level. In order to determine whether there is an overrepresentation of specific disabilities that are tested out of level, it is important for policymakers, educators, and parents to know out-of-level test prevalence data by disability category. This type of data-based information can drive policy, instructional, and assessment decisions so that more students with disabilities are better supported in reaching high learning expectations. However, these decisions are impossible to make when state and district level large-scale assessment results are reported in such a way that the public does not know where and how out-of-level test scores are reported.

Issue #4—Reporting out-of-level test scores is viewed as a statistical problem.

Some states identified specific statistical problems in reporting scores for below grade level testing. Interviewees expressed concerns about CRT scores that cannot be used for transforming out-of-level to on-level scores. In particular, those interviewed from Mississippi were concerned about the misleading nature of entering all out-of-level test scores in the lowest proficiency level for accounting progress toward grade level content standards. An interviewee from Hawaii acknowledged that the omission of out-of-level test scores to avoid confidentiality violations could potentially skew reporting on local data.

While we acknowledge that these are justifiable concerns, we raise an additional statistical concern that is also critical to suitable large-scale assessment and accountability programs. Generally speaking, when mathematically manipulating large numbers, the omission of a few test scores, as in the case of reporting state level large-scale assessment results, does not affect the numeric outcomes. However, ignoring the test scores of even a few students with disabilities falls short of current pressures to ensure that *all* students achieve grade-level content standards. Current federal mandates have moved the field beyond simply focusing on the statistics of reporting to the public. Instead, educators and policymakers alike are challenged to think critically about

improving classroom instruction that in turn will augment the statistics of states' reported test results.

Remaining Challenges

This report has described states' reporting practices at specific points in time in order to flesh out the many issues related to reporting the results of out-of-level tests that are used for statewide testing. In doing so, our intent has not been to ascribe fault to state level personnel for not having resolved state-specific problems in their reporting practices. Since states have come under extreme scrutiny to demonstrate improved student results as measured by statewide testing over the past decade, it is especially important to acknowledge that these reporting practices on which we collected data were devised prior to the enactment of NCLB in 2001.

To that end, we conclude this report by identifying remaining challenges that constrain the research community as we respond to the informational needs of those who make decisions about out-of-level testing reporting practices.

- *There is a lack of consistency in states' out-of-level testing policies, which impedes the development of recommended guidelines for reporting test data.* The complexity of parsing out useful, data-based information to guide test administration and the subsequent public reporting is further compounded by the variety of contexts in which students with disabilities are tested out of level.
- *Quite often, SEAs lack communication among different divisions, all of which are important to the successful implementation of large-scale assessment and accountability policy.* States vary as to whether out-of-level testing is a special education or an assessment issue. States may contract out to other agencies that do not share common space further impeding collaboration. High demands on SEA personnel professional time allows for little cross-disciplinary communication. The political ramifications of testing students with disabilities out of level hamper the open sharing of information. Each of these factors also constrains the development and implementation of high-quality research processes.
- *There are various factors inherent to states' out-of-level testing policy that constrain reporting practices.* State policies define out-of-level testing in various ways that dictate how scores can be reported (e.g., modifications are used differently from accommodations). Some students' test scores are deleted from accountability indices to avoid violating confidentiality regulations. Again, the type of instrument administered out of level determines how scores can be treated mathematically for accountability purposes. Research can identify and describe these factors including the related policy constraints. However, when these constraints are

played out within politically charged environments, conducting educational research becomes increasingly complicated.

In sum, fully reporting out-of-level test scores is a necessary step toward understanding better the needs that out-of-level testing is said to meet. To foster the acquisition of grade-level standards for all students, it is necessary for the educational community to re-focus its attention to the learning needs of students with disabilities. In doing so, public reporting becomes more than the numbers inherent in the accountability process. Reporting practices can be a tool to support students with disabilities as they strive for improved learning outcomes. Findings from this research study support the Federal decision that out-of-level testing is not an acceptable means for fulfilling the state's assessment requirements under NCLB (Federal Register, July 5, 2002, pp. 45044-45).

References

Bielinski, J., Minnema, J., Thurlow, M., & Guven, K. (2003). *Testing students with disabilities out of level: State prevalence and performance results* (Out-of-Level Testing Report 9). University of Minnesota, MN. University of Minnesota, National Center on Educational Outcomes.

Bielinski, J., Thurlow, M., Callender, S., & Bolt, S. (2001). *On the road to accountability: Reporting outcomes for students with disabilities*. (Technical Report 32). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Federal Register (July 5, 2002). Title I -- Improving the Academic Achievement of the Disadvantaged, Volume 67 (129). Retrieved April 29, 2003 from <http://www.ed.gov/legislation/FedRegister/finrule/2002-3/070502a.html>

Minnema, J., Thurlow, M., & Scott, J. (2001). *Testing students out of level in large-scale assessments: What states perceive and believe*. (Out-of-Level Testing Report 5). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M., & Minnema, J. (2001). *States' out-of-level testing policies* (Out-of-Level Testing Report 4). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M., Wiley, H. I., & Bielinski, J. (2003). *Going public: What 2000-2001 reports tell us about the performance of students with disabilities* (Technical Report 35). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.



The College of Education
& Human Development

UNIVERSITY OF MINNESOTA

NCEO is an affiliated center of the Institute on Community Integration