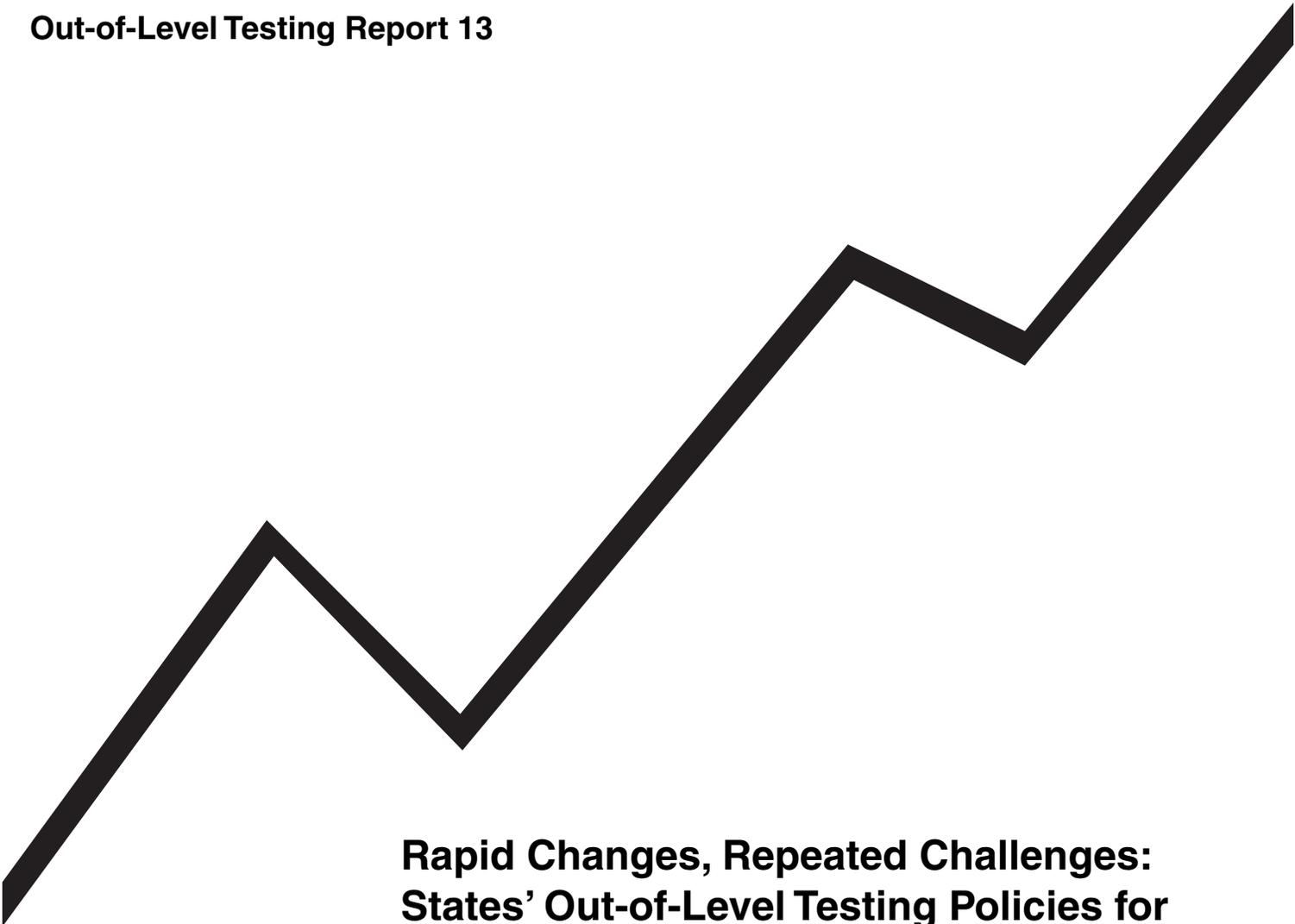


NATIONAL
CENTER ON
EDUCATIONAL
OUTCOMES

This document has been archived by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it contains may be out of date.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>



**Rapid Changes, Repeated Challenges:
States' Out-of-Level Testing Policies for
2003-2004**



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 13

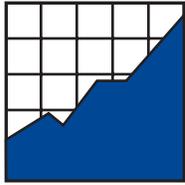
Rapid Changes, Repeated Challenges: States' Out-of-Level Testing Policies for 2003-2004

Gretchen VanGetson • Jane Minnema • Martha Thurlow

September 2004

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

VanGetson, G., Minnema, J., & Thurlow, M. (2004). *Rapid changes, repeated challenges: States' out-of-level testing policies for 2003-2004* (Out-of-Level Testing Report 13). 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
Ann T. Clapper	Michael L. Moore
Christopher J. Johnstone	Rachel F. Quenemoen
Jane L. Krentz	Dorene L. Scott
Sheryl Lazarus	Sandra J. Thompson
Kristi K. Liu	Martha L. Thurlow, Director
Jane E. Minnema	

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://nceo.info>

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

The purpose of this research study was to illustrate the ways in which out-of-level testing policies have changed over the three years from 2000-2001 to 2003-2004. In 2000-2001, 12 states were using out-of-level tests to measure student progress toward content standard proficiency. A detailed research study conducted at the end of 2000-2001 to examine the out-of-level testing policies of these 12 states (Thurlow & Minnema, 2001) revealed that policies varied considerably across states. Much has happened in the years since 2000-2001. By 2003-2004, the number of states having a version of below grade level testing (out-of-level or levels testing) as an option in their large-scale assessment programs had increased to 17 states.

We conducted this study by comparing the results of the first study (Thurlow & Minnema, 2001) with the results of our 2003-2004 policy review. For the 2003-2004 review, we extracted thematic results to compare with the thematic results from our first out-of-level policy review.

Our comparison of themes from the past, present, and future of out-of-level testing provided six summative discussion points about the condition of out-of-level testing in this decade:

1. It was discovered that recent federal legislation is not reflected in states' use of out-of-level or levels testing.
2. States were found to use a greater variety of out-of-level or levels testing classification terms than they did in 2000-2001.
3. Some states had changed the qualifications for students allowed to participate in out-of-level or levels testing.
4. Some states had added more content areas assessed in out-of-level or levels tests.
5. There has been an increase in state-level reporting of out-of-level or levels testing scores although states have tended to aggregate these results with on-level scores.
6. The long term effects of using out-of-level or levels tests remain unknown despite the increased use of high-stakes assessments.

The continuing controversy surrounding out-of-level and levels testing will remain of interest to practitioners, policymakers, and researchers at all levels of the American educational system.

Overview

Since the enactment of the No Child Left Behind (NCLB) Act of 2001, states have endeavored to include *all* subgroups of students in standards-based assessment that measure academic progress toward proficiency. In the past, states have allowed various forms of testing, such as out-of-level testing, as a means to include more students in statewide testing. Out-of-level testing, most often thought of as the administration of a large-scale assessment above or below the grade in which a student is enrolled in school, has allowed states to boost their participation rates on either a state or district basis. Given the federal mandate that *all* students must receive challenging, grade-level curriculum to support their acquisition of grade-level content standards, states have been forced to look critically at their large-scale assessment policies and the local effects of implementing those policies.

States have responded to NCLB by altering their large-scale assessment programs. NCLB requires states to include all students in their large-scale statewide assessments aligned to grade-level achievement standards, and NCLB regulations only allow for alternate achievement standards for children with the most significant cognitive disabilities, and limit the percent of students who can demonstrate proficiency on alternate achievement standards to one percent of the student population unless an exception is obtained (Federal Register, 2003). Yet there is still a cohort of students not apparently achieving on-grade level who may be receiving instruction on content standards below their grade of enrollment. Some states have developed out-of-level testing options in their large-scale statewide assessment programs. As large-scale assessment policies have shifted over the years, so too have states' out-of-level testing policies.

In 2000-2001, 12 states were using out-of-level tests to measure student progress toward content standard proficiency. We conducted a study to examine the out-of-level testing policies of these 12 states from the 2000-2001 school year (Thurlow & Minnema, 2001) and found that policies varied considerably from state to state. In 2003-2004, we conducted another study and found the number of states that reported having a version of below grade level testing as an option in their large-scale assessment programs increased to 17. Some of the 12 states from 2000-2001 were still included in the 17 states from 2003-2004. Other states had since eliminated out-of-level testing while new states had added this approach to testing. All states had revised their out-of-level testing policy by making either major or minor changes.

The purpose of the current study was to examine the ways in which out-of-level testing policies have changed over the past three years. We accomplish this goal by comparing the results of the 2000-2001 out-of-level testing study (Thurlow & Minnema, 2001) with the results of our 2003-2004 policy review. For our comparison, we describe out-of-level testing policies that clarify the status of out-of-level testing as well as anticipated changes in states' policies in the future.

Two research questions were addressed in this study:

- (1) How have states' out-of-level testing policies changed from school year 2000-2001 to school year 2003-2004?
- (2) Which states have added or discontinued out-of-level testing since 2000-2001?

Method

Using multiple sources of data, we reviewed states' online or paper versions of out-of-level testing policies to glean relevant policy information for those states that tested students with disabilities in large-scale assessment programs below grade level during school year 2003-2004. To begin our study, we used a four step process for organizing our set of out-of-level testing policies for review. First, we began with our out-of-level testing policy files that we maintain at NCEO. Since we have updated this file of policies annually over the past three years, we had relatively current information from which to initiate the study. Next we revisited our results from 2000-2001 to compare those data to our current policy information on file. Third, to determine whether any states had begun testing out of level since 2001, we examined data from the 2003 Survey of State Directors of Special Education (Thompson & Thurlow, 2003). In this survey, states were asked to report on any below grade level testing in their large-scale assessment programs. With this information, we updated our list to 17 states that were possibly testing out of level.

To begin our data collection, we searched the state education agency Web site for each of the 17 states that were identified as testing out of level in 2003-2004. All 17 states had some policy information available online that was related to out-of-level testing. That information was downloaded and printed for our files. If we were unable to locate in-depth out-of-level testing policy information online, we contacted states directly to request paper copies of their most recent policies.

Similar to the method used in the 2000-2001 out-of-level testing study (Thurlow & Minnema, 2001), we reviewed each state's out-of-level testing policy individually to determine the specific content of the policy on a state by state basis. Then, to understand how policies had changed over time, we considered all of the policies as a composite data set from which we identified state-specific contextual features, the current status, and significant content details of states' out-of-level testing policies. We charted this composite data set of policy information into tables to further highlight state to state comparisons. An individual from each state was invited to review his or her states' information for accuracy prior to inclusion in the data tables. Finally, we examined the data set holistically to extract thematic results to compare with the thematic results from the 2000-2001 out-of-level policy review study. Our comparison of overarching

themes from the past, present, and future of out-of-level testing provided summative discussion points about the condition of out-of-level testing.

Status of Out-of-Level Testing in States

The progression of states' use of out-of-level testing since the 2000-2001 study is shown in Table 1. In 2001-2002, the year following the first study, three states (Hawaii, Oregon, and Texas) added or acknowledged the existence of an out-of-level testing option to their large-scale state-wide assessment programs. Since then, three additional states (Nebraska, North Carolina, and Tennessee) developed an out-of-level testing option. During the same time frame, many states eliminated this option from their assessment programs. Two states (Alaska and North Dakota) discontinued testing out of level in 2001-2002. Another state (West Virginia) discontinued out-of-level testing in 2002-2003. Four more states (Connecticut, Delaware, Hawaii, and Louisiana) eliminated out-of-level testing in 2003-2004. There were six states (Arizona, California, Iowa, South Carolina, Utah, and Vermont) that maintained out-of-level testing across all of the years following the first study. Many did so by revising the policy content—with considerable change in some cases.

Table 2 expands on Table 1 by identifying states that have discontinued their out-of-level testing option in the past and those states that anticipated discontinuing or changing out-of-level testing in the near future. For instance, two states (Tennessee and Utah) plan on discontinuing out-of-level testing in 2004-2005. Further, four states (Arizona, California, Texas, and Vermont) anticipated future changes to their out-of-level testing policies, but were unsure about the details of those changes at this time.

Out-of-Level Versus Levels Testing

In the 2003 Survey of State Directors of Special Education (Thompson & Thurlow, 2003), respondents were asked to answer the following question: “Does your state currently have out-of-level or levels testing options?” Based on the language used to respond to this survey item, we separated those states that indicated using out-of-level testing from those that indicated using a levels testing option. Table 3 contains this information. There were 14 states (Arizona, California, Connecticut, Delaware, Hawaii, Iowa, Louisiana, Mississippi, Nebraska, North Carolina, South Carolina, Tennessee, Utah, and Vermont) that indicated that they offered an out-of-level testing option in 2003. Three states (Kansas, Oregon, and Texas) indicated that they offered a levels testing option in 2003. While both out-of-level testing and levels testing assess

Table 1. Implementation and Discontinuation History of Out-of-Level Testing in States

State	Prior to 1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004
Alaska		X	X			
Arizona	X	X	X	X	X	X
California	X	X	X	X	X	X
Connecticut	X	X	X	X	X	
Delaware			X	X	X	
Hawaii				X	X	
Iowa	X	X	X	X	X	X
Kansas	Unknown	Unknown	Unknown	Unknown	X	X
Louisiana		X	X	X	X	
Mississippi			X	X	X	X
Nebraska						X
North Dakota	X	X	X			
North Carolina					X	X
Oregon				X	X	X
South Carolina		X	X	X	X	X
Tennessee					X	X
Texas				X	X	X
Utah	X	X	X	X	X	X
Vermont	X	X	X	X	X	X
West Virginia	X	X	X	X		

Table 2. Recent and Future Discontinuations of Out-of-Level Testing

Discontinued 2001-2002	Discontinued 2002-2003	Discontinued 2003-2004	Will Discontinue 2004-2005	Anticipate Future Changes
Alaska	West Virginia	Connecticut	Tennessee	Arizona
Alabama		Delaware	Utah	California
Georgia		Hawaii		Texas
North Dakota		Louisiana		Vermont

students below the grade in which they are enrolled in school, we make the distinction between these approaches throughout this report. Some states prefer the levels approach to testing for assessing academic proficiency because the test levels are created on a common scale so that scores from below grade level and grade of enrollment tests can be compared.

Table 3. States Using Out-of-Level or Levels Testing in 2003

Out-of-Level Testing	Levels Testing
Arizona	Kansas
California	Oregon
Connecticut*	Texas
Delaware*	
Hawaii*	
Iowa	
Louisiana*	
Mississippi	
Nebraska	
North Carolina	
South Carolina	
Tennessee	
Utah	
Vermont	

* Indicates that the state discontinued testing out-of-level in 2003-2004.

Out-of-Level Testing Context

As out-of-level testing has changed in the recent past, so has the context within which out-of-level testing is implemented. To fully understand out-of-level testing in a state, it is helpful to first understand the state's assessment context. Tables 4 and 5 present aspects of regular large-scale assessments in states that used out-of-level (Table 4) or levels (Table 5) testing during the 2003-2004 school year. In each table, we report for each state the name of the state's regular assessment program (if the program has a name), the tests included in the state's assessment program, the grade levels assessed by each test, and the content areas tested.

Generally speaking, there was considerable variability across states in the tests used for assessing standards-based academic proficiency. For example, some states used a combination of criterion-referenced and norm-referenced tests while other states used only one of these assessment types. Additionally, some states employed one test to assess all content areas and grade levels, while others used a combination of tests that assessed different content areas and grade levels. There was also wide variability in the grade levels tested by the states' assessments. These state assessments spanned all grades, from kindergarten through 12th grade, depending on the state, the test administered, and the content area. The content areas tested were more consistent across states, covering reading, writing, and mathematics, and less often, social studies and science. Yet no two states tested the same content areas in the same manner, resulting in more differences than similarities across states.

Table 4. Out-of-Level Testing Context - State Assessments by States

State	State Testing Program Name	State Tests	Grades Tested	Content Areas Tested
Arizona	(no testing program name)	AZ Instrument to Measure Standards (AIMS)	3 rd , 5 th , 8 th , high school	Reading, writing, math
		Stanford 9 Achievement Test	3 rd – 11 th	Test battery
California	Standardized Testing & Reporting Program (STAR)	California Standards Tests (CST)	2 nd – 11 th	English/language arts, math, writing, social science, science
		California Achievement Tests, Sixth Edition (CAT-6)	2 nd – 11 th	Reading/language, spelling, math
		Spanish Assessment of Basic Education, Second Edition (SABE-2)	2 nd – 11 th	Reading, spelling, language, math
		California High School Exit Examination (CAHSEE)	10 th	Language arts, math
Connecticut*	(no testing program name)	Connecticut Mastery Test (CMT)	4 th , 6 th and 8 th	Reading, writing, math
		Connecticut Academic Performance Test (CAPT)	10 th	Math, reading, writing, science
Delaware*	(program name same as test name)	Delaware Student Testing Program (DSTP)	2 nd – 10 th	English/language arts, math, social studies, science, writing
Hawaii*	(program name same as test name)	Hawaii Content and Performance Standards (HCPS II) State Assessment	3 rd , 5 th , 8 th and 10 th	Reading, writing, math
Iowa	No statewide assessment program	Iowa Tests of Basic Skills (ITBS)	3 rd – 8 th (4 th and 8 th required)	Minimum: Reading comprehension, math concepts and estimation, math problem solving and data interpretation, science
		Iowa Tests of Educational Development (ITED)	9 th – 12 th (11 th required)	Minimum: Reading comprehension, math concepts and problem solving, science

State	State Testing Program Name	State Tests	Grades Tested	Content Areas Tested
Louisiana*	Louisiana Criterion-Referenced Testing Program	Louisiana Educational Assessment Program for the 21 st Century (LEAP 21)	4 th and 8 th	English/language arts, math, science, social studies
		Graduation Exit Examination for the 21 st Century (GEE 21)	10 th and 11 th	English/language arts, math, science, social studies
	Louisiana Statewide Norm-Referenced Testing Program (LSNRTP)	Iowa Tests of Basic Skills (ITBS)	3 rd , 5 th , 6 th , and 7 th	Test battery
		Iowa Tests of Educational Development (ITED)	9 th	Test battery
Mississippi	Mississippi Grade Level Assessment Program	Mississippi Curriculum Test	2 nd – 8 th	Reading, language, math
		Writing Assessment	4 th and 7 th	Writing
		TerraNova	6 th	Reading/language arts, math
Nebraska	School-based Teacher-led Assessment and Reporting System (STARS)	Statewide Writing Assessment	4 th , 8 th and 11 th	Writing
		STARS Reading	4 th , 8 th , and 11 th	Reading
		STARS Math	4 th , 8 th , and 11 th	Math
North Carolina	(program name same as test name)	North Carolina Testing Program	3 rd – 8 th , 10 th	Reading, math
South Carolina	(no testing program name)	Palmetto Achievement Challenge Tests (PACT)	1 st – 8 th (1 st , 2 nd grade optional)	English/language arts, math, science, social studies
		High School Assessment Program (HSAP)	10 th	English/language arts, math
Tennessee	Tennessee Comprehensive Assessment Program (TCAP)	Achievement Test	3 rd – 8 th (K – 2 nd optional)	Reading, language arts, math, science, social studies
		Gateway Testing Initiative	High school (beginning with 9 th grade students in 2001-2002)	Math, science, language arts
		Writing Assessment	5 th , 8 th , and 11 th	Writing

State	State Testing Program Name	State Tests	Grades Tested	Content Areas Tested
Utah	Utah Performance Assessment System for Students (U-PASS)	Core Assessment Criterion-Referenced Tests	1 st – 11 th	Reading/language arts, math, science
		Direct Writing Assessment	6 th and 9 th	Writing
		Stanford Achievement Test, 9 th Edition	3 rd , 5 th , 8 th , and 11 th	Reading/language arts, math, science, social studies
		Utah Basic Skills Competency Test	10 th	Reading, writing, math
Vermont	Vermont Comprehensive Assessment System (CAS)	Vermont Developmental Reading Assessment (DRA)	2 nd	Reading
		New Standards Reference Exams (NSRE)	4 th , 8 th , and 10 th	English/language arts, math
		VT-PASS	5 th , 9 th , and 11 th	Science

* Indicates that the state discontinued testing out-of-level in 2003-2004.

Table 5. Levels Testing Context - State Assessments by States

State	State Testing Program Name	State Tests	Grades Tested	Content Areas Tested
Kansas	Kansas State Assessments	Reading Assessment	5 th , 8 th , and 11 th	Reading
		Mathematics Assessment	4 th , 7 th , and 10 th	Math
		Science Assessment	4 th , 7 th , and 10 th	Science
		Social Studies Assessment	6 th , 8 th , and 11 th	Social Studies
Oregon	Oregon Statewide Assessment	Knowledge and Skills Assessments	3 rd - 8 th , 10 th	Reading/literature, math, science (starting in 5 th grade), social science (starting in 5 th grade)
		Performance Assessments	5 th , 8 th , and 10 th	Math problem solving, writing
Texas	(no testing program name)	Texas Assessment of Knowledge and Skills (TAKS)	3 rd – 11 th	Reading (3 rd – 9 th), writing (4 th and 7 th), English language arts (10 th and 11 th), math, science (5 th , 10 th , and 11 th), social studies (8 th , 10 th , and 11 th)
		Texas Assessment of Academic Skills (TAAS)	High School	Reading, math, writing

Out-of-Level Testing Policy Content

Out-of-level testing policies provide state-level guidance for local-level, out-of-level, or levels testing implementation. The policy language helps to ensure consistent implementation throughout the state. As with states' regular assessment programs, the policy content in out-of-level or levels testing policies differed widely. To more clearly describe policy information across states, we have separated policy language into three categories: state-level policy features, instrument characteristics, and test score use. In describing this policy information, we gleaned thematic generalizations from reviewing each table.

State-Level Policy Features

Tables 6 and 7 highlight important features of each state's out-of-level or levels testing policies respectively. These features include the name of the written document that included the policy, the out-of-level or levels testing classification, the inclusion of selection criteria within the policy, and the students eligible for this testing option. The themes of policy features that emerged from the 2003-2004 policies were the same as those that emerged in 2000-2001.

State level policies on out-of-level testing were in a variety of written formats. Some states included out-of-level or levels testing policy information in their test administration information (Arizona, Iowa, and Oregon), and some states included this information in their test participation guidelines (Connecticut, Delaware, Hawaii, Mississippi, South Carolina, and Utah). Three states (Tennessee, Kansas, and Texas) had a special document devoted to the out-of-level or levels test that included policy information. Two states (North Carolina and Nebraska) included policy information in their large-scale assessment updates, and another state (Vermont) included policy information directly in the form that practitioners use to document participation. For example, the form includes checkboxes for practitioners to indicate which regular grade level assessment the out-of-level test will replace, the allowable out-of-level grade level assessment that the student should take, and required procedures that helped guide this particular decision. Also, the criteria for participation in an out-of-level test are included in the description of out-of-level testing on the form.

One state (Louisiana) used an assessment interpretive guide to disseminate policy information while another state (California) included this information in its accountability workbook. It should be noted that what appeared to be differences in policy formats or document names may only be language differences in the states. For instance, an administration manual may have contained the same information as participation guidelines, but simply presented under a different name. Nevertheless, the variability in language and, subsequently, policy formats, is important to consider because it is integral to locating a state's out-of-level or levels testing policy information.

States that allowed out-of-level or levels testing did not treat these testing options similarly in their written policies. There were many labels that states used to classify out-of-level or levels testing. Six states called these options alternate assessments (Arizona, Connecticut, Louisiana, North Carolina, Tennessee, and Texas), which was the most common term. Two states (Delaware and Hawaii) used the term accommodation while two other states (South Carolina and Utah) used the term modification. Two more states (Mississippi and Nebraska) referred to out-of-level testing as instructional level testing, but Nebraska also used the term below grade testing along with another state (California). Finally, four states used a classification that was exclusive to the state. These classifications included modified assessment (Kansas), challenging another benchmark (Oregon), adapted assessment (Vermont), and on-level testing because out-of-level testing is considered to be the same thing (Iowa).

Most states provided criteria in their assessment policies for selecting students for out-of-level or levels testing. The majority of the states that administered out-of-level or levels testing established some form of selection criteria for student eligibility for these testing options. Four of these states (California, Delaware, Mississippi, and Texas) further limited these criteria by placing grade level restrictions on out-of-level or levels testing participation. One state (Iowa) did not specify state-level selection criteria because this state maintained out-of-level testing participation as a local level decision. Finally, two states (Nebraska and North Carolina) did not include specific selection criteria in their written policies, which were in the form of large-scale assessment briefs or updates. Both documents were written generally without specific detail about below grade level testing in their states.

Students with disabilities who had Individualized Education Programs (IEPs) were typically the only students who could be tested out of level or with a levels test. Eleven states (Arizona, Connecticut, Delaware, Hawaii, Louisiana, Mississippi, Nebraska, South Carolina, Tennessee, Utah, and Texas) required that a student must have an IEP to be considered for out-of-level or levels testing. Of these states, some had criteria beyond having an IEP. For example, Arizona required that the student must have a significant cognitive disability. Four states (California, Kansas, North Carolina, and Oregon) required that a student must have either an IEP or a 504 plan to be considered for out-of-level or levels testing. One state (Iowa) made this a local level decision, and another state (Vermont) allowed any student recommended to the state by the Student Support Team as being eligible for out-of-level testing. No student in Vermont could be tested below grade level without specific state-level approval.

Table 6. Out-of-Level Testing Policies by States - State-Level Policy Features

State	Written Policy Format	Out-of-Level Classification	Selection Criteria	Students Tested
Arizona	Administration of AIMS and SAT9 to Students with Disabilities	Alternate Assessment- together with the AIMS-A	Yes	Students with IEPs who are labeled as significantly cognitively disabled
California	Attachment F: Accountability Workbook	Below Level Testing	Yes; only available for 5 th grade students and above	Students with IEPs or 504 Plans
Connecticut*	Assessment Guidelines (9 th ed.)	Alternate Assessment	Yes	Students with IEPs
Delaware*	Delaware Student Testing Program: Guidelines for the Inclusion of Students with Disabilities and Students with Limited English Proficiency	Accommodation	Yes; only available to 5 th , 8 th , and 10 th grade students	Students with IEPs
Hawaii*	HCPS II State Assessment Student Participation Information	Accommodation	Yes	Students with IEPs
Iowa	Policy and guidance included in Directions for Administration	Considered to be the same as on-level testing	Locally determined	Local decision
Louisiana*	Louisiana Statewide Norm-Referenced Testing Program 2003 Interpretive Guide	Alternate Assessment	Yes	Students with IEPs
Mississippi	Mississippi Statewide Assessment System: Guidelines for Student with Disabilities and English Language Learners	Instructional Level Testing	Yes; only available for students in 2 nd – 8 th grades	Students with IEPs, if recommended by the IEP team
Nebraska	STARS Update	Below Grade/ Instructional Level Testing	Not specified in policy	Students with IEPs
North Carolina	Assessment Brief: North Carolina Alternate Assessment Academic Inventory	Alternate Assessment	Not specified in policy	Students with IEPs or 504 Plans
South Carolina	Testing Students with Disabilities: Guidelines for IEP Teams	Modification	Yes	Students with IEPs
Tennessee	Tennessee Alternate Portfolio Assessment	Alternate Assessment	Yes	Students with IEPs who meet additional criteria

State	Written Policy Format	Out-of-Level Classification	Selection Criteria	Students Tested
Utah	Requirement for Participation of Utah Students with Special Needs in the Utah Performance Assessment System for Students (U-PASS)	Modification	Yes	Students with IEPs
Vermont	Vermont Statewide Assessment System: Documentation of Eligibility for Alternate Assessment	Adapted Assessment	Yes; must be approved for use in accountability by the Department of Education	Students with IEPs, 504 Plans, or a recommendation by the Student Support Team

* Indicates that the state discontinued testing out-of-level in 2003-2004.

Table 7. Levels Testing Policies by States - State-Level Policy Features

State	Written Policy Format	Level Testing Classification	Selection Criteria	Students Tested
Kansas	Kansas Modified Assessments: Eligibility Criteria and Overview for 2003-2004 Academic Year	Modified Assessment	Yes	Students with IEPs or 504 Plans
Oregon	Knowledge and Skills Administration Manual	Challenging Another Benchmark	Yes	Students with IEPs or 504 Plans
Texas	State-Developed Alternative Assessment (SDAA): Information Brochure Revised	Alternate Assessment	Yes; only available for students in 3 rd -8 th grades	Students with IEPs

Instrument Characteristics

The characteristics of the assessment instruments used in out-of-level or levels tests are presented in Tables 8 and 9, respectively. Three general themes were derived from these descriptive data.

Both criterion-referenced and norm-referenced tests were used for out-of-level tests. The type of test used in states' out-of-level testing options varied. Eight states (Connecticut, Mississippi, Nebraska, North Carolina, South Carolina, Tennessee, Utah, and Vermont) used only a criterion-referenced test for their out-of-level assessment. Four states (Arizona, California,

Delaware, and Hawaii) administered a combination of criterion-referenced and norm-referenced assessments. Two states (Iowa and Louisiana) used only a norm-referenced test for out-of-level testing purposes. Louisiana's use of a norm-referenced test in out-of-level testing was unique in that its general assessment included both criterion-referenced (LEAP 21; GEE 21) and norm-referenced (ITBS; ITED) components, but students who took the assessment out of level took only an extended version of the norm-referenced component in lieu of the criterion-referenced test.

There was wide variability in the allowed test grade levels in out-of-level or levels tests. Wide variability was the only way to summarize the grade levels at which states allowed students to be tested by out-of-level or levels tests. Four states (Arizona, Hawaii, Mississippi, and Nebraska) had the least restrictive guidelines in that a student enrolled in any grade level could take any available test level offered as long as the test level was administered at the student's instructional level. Five other states (North Carolina, Oregon, South Carolina, Texas, and Utah) allowed any test level that matched the student's instructional level within certain grade level limits. For example, North Carolina allowed any grade 3 through 8 test level to be administered out of level as long as the test grade level matched the student's instructional level. In other words, out-of-level tests administered below grade 3 were not allowed. Two states (Delaware and Vermont) allowed out-of-level test presentations only at the grade levels of the general assessment. For instance, students in Delaware in grades 5, 8, and 10 were restricted to taking out-of-level tests at grades 3, 5, or 8, which were the grade levels at which the general standards-based measure was administered. Finally, four states (California, Connecticut, Iowa, and Louisiana) set a limit on the number of levels below which a test could be administered out of level. California allowed no more than two test grade levels below the student's grade of enrollment and Iowa allowed no more than two to four test grade levels below the student's grade of enrollment. Connecticut and Louisiana allowed no more than three test grade levels below the student's grade or enrollment, and allowed the student to take the out-of-level test at more than one test grade level (i.e., grade 3 in reading and grade 5 in math, based on the student's academic ability).

All states tested core content areas in out-of-level or levels testing. The core content areas of reading/language arts and math were assessed in all states that used out-of-level or levels testing. Eight states (Arizona, California, Connecticut, Delaware, Hawaii, Oregon, Texas, and Utah) also included writing tests, 11 states (California, Connecticut, Delaware, Iowa, Kansas, Louisiana, Nebraska, Oregon, South Carolina, Tennessee, and Utah) included science tests, and nine states (California, Delaware, Kansas, Louisiana, Nebraska, Oregon, South Carolina, Tennessee, and Utah) included social studies tests in their out-of-level or levels testing policies. One state (Nebraska) also assessed listening and speaking skills out of level. Four states (Arizona, Connecticut, Delaware, and Hawaii) allowed students to test only one content area out of level, while one state (California) required students tested out of level to take all content areas at the same test grade level.

Table 8. Out-of-Level Testing Policies - Instrument Characteristics by States

State	Type of Test	Grade Levels Tested Out	Content Areas Tested
Arizona	NRT/CRT	Any available level to match test level to instructional level	Reading, writing, and math (May test one area)
California	NRT/CRT	No more than two levels below grade of enrollment	Reading, language arts, writing, math, science, and social studies (Must take all tests offered at test level)
Connecticut*	CRT	No more than three test levels below grade of enrollment (May take test at different levels)	Math, reading, writing, and science (May test one area)
Delaware*	NRT/CRT	Only available test grade levels were grades 3, 5, and 8	English/language arts, math, social studies, science, and writing (May test one area)
Hawaii*	NRT/CRT	Any available level to match test level to instructional level	Reading/writing and math (May test one area)
Iowa	NRT battery	2-4 grade levels below grade of enrollment for grades 3 - 12	Minimum: Reading comprehension, math concepts and problem solving, and science
Louisiana*	NRT (in lieu of CRT)	At least 3 grade levels below grade of enrollment for English/language arts or math; may test two different test levels	English/language arts, math, science, and social studies
Mississippi	CRT	Any available level to match test level to instructional level	Reading, language, and math
Nebraska	CRT	Any available level to match test level to instructional level	Reading, math, science, social studies, listening, and speaking
North Carolina	CRT	Any available level for grades 3 - 8	Reading and math
South Carolina	CRT	Grades 1 – 8 to match instructional level	English/language arts, math, science, and social studies
Tennessee	CRT	Unknown	English, language arts, math, social studies, and science
Utah	CRT	Any available level in grades 1 – 11 to match test level to instructional level; usually at least 3 levels below grade of enrollment	Reading/language arts, math, science, writing, and social studies
Vermont	CRT	Grades 4, 8, and 10 (New Standard Reference Exam levels offered)	English/language arts and math

* Indicates that the state discontinued testing out-of-level in 2003-2004.

Table 9. Levels Testing Policies - Instrument Characteristics by States

State	Type of Test	Grade Levels Tested Out	Content Areas Tested
Kansas	Unknown	Unknown	Math, reading, science, and social studies
Oregon	CRT	Any available level for grades 3 – 8 to match test level to instructional level	Reading/literature, math, science, social science, and writing
Texas	CRT	Any available level of test (K – 8) to match test level to instructional level	Reading, math, and writing

Test Score Use

Tables 10 and 11 provide information on how states use out-of-level (Table 10) or levels (Table 11) test scores. Three themes emerged from this set of data.

Most states did not equate out-of-level or levels test scores to on-level test scores. Twelve states (California, Connecticut, Delaware, Hawaii, Louisiana, Mississippi, Nebraska, North Carolina, Oregon, South Carolina, Texas, and Utah) responded that they do not attempt to equate out-of-level or levels test scores to on-level scores. One state (Arizona) indicated that this process is in development so that it could not provide a definitive answer. Two states (Iowa and Vermont) answered that they do equate out-of-level test scores to on-level test scores. Iowa used a standard developmental growth scale and Vermont used score transformation rules to create this linkage.

States used a variety of state level reporting methods. Across the 17 states that used out-of-level or levels testing in 2003-2004, eight different state level reporting methods were used. Some states (Delaware, Iowa, and Louisiana) aggregated these scores at the student’s grade of enrollment. Other states (South Carolina, Texas, and Utah) disaggregated these data; for example, Texas reported results from the State-Developed Alternative Assessment (SDAA), the name for its levels test, separately from any other assessment. Some states chose to report at the test grade level instead of the student’s enrolled grade level, with one state (Mississippi) aggregating out-of-level test scores and another state (Connecticut) reporting out-of-level test scores in an unspecified manner. One state (Arizona) only reported out-of-level test scores online and another state (North Carolina) only reported a summary of these data. Additionally, one state (Vermont) only reported out-of-level test scores within the adequate yearly progress (AYP) index for that state by converting those scores using a point scale. Two states (California and Nebraska) indicated that they do not report out-of-level test scores at the state level.

Most states included out-of-level testing scores in the lowest proficiency level for account-

ability reporting purposes. The most common accountability reporting practice for states testing out of level or using levels testing was to include those students’ scores at the lowest proficiency level at the student’s grade of enrollment. Nine states (California, Connecticut, Delaware, Hawaii, Iowa, Louisiana, Nebraska, Utah, and Vermont) used this reporting procedure. Three states (Arizona, Mississippi, and North Carolina) indicated that they included out-of-level test scores in the one percent allowance cap for alternate assessments according to current U.S. Department of Education regulations. But, North Carolina noted that any overflow beyond the one percent cap was included at the lowest proficiency level at the student’s grade of enrollment. Two states (South Carolina and Tennessee) included these scores at the score-appropriate proficiency level. But, South Carolina included these scores at the test grade level while Tennessee included these scores at the student’s grade of enrollment.

Table 10. Out-of-Level Testing Policies – Test Score Use by States

State	Equated to On-level Scores	State Level Reporting Methods	Accountability Reporting
Arizona	In development	Scores are reported online; not included in state accountability system AZLEARNS	Because out-of-level testing is an alternate assessment, results are included in the 1% cap for proficiency levels
California	No	Nonstandard scores not reported at state level	Included at the lowest proficiency level at the grade of enrollment
Connecticut*	No	Reported for grade level of test	Included at the lowest proficiency level at the grade of enrollment
Delaware*	No	Aggregated at grade of enrollment	Included at the lowest proficiency level at the grade of enrollment
Hawaii*	No	Disaggregated at grade of enrollment	Included at the lowest proficiency level at the grade of enrollment
Iowa	Yes (Standard developmental growth scale)	Aggregated at grade of enrollment	Included at the lowest proficiency level at the grade of enrollment
Louisiana*	No	Aggregated at grade of enrollment	Included at the lowest proficiency level at the grade of enrollment
Mississippi	No, and reported separately from grade level testing except for AYP	Aggregated at test level	Follows current USDE regulation/guidance
Nebraska	No	Not reported	Included at the lowest proficiency level at the grade of enrollment

State	Equated to On-level Scores	State Level Reporting Methods	Accountability Reporting
North Carolina	No	Summary of data only	Included at the lowest proficiency level at the grade of enrollment beyond the 1% allowance cap
South Carolina	No	Disaggregated	Included in appropriate proficiency level at the test grade level
Tennessee	Unknown	Unknown	Included in appropriate proficiency level at the grade of enrollment
Utah	No	Disaggregated	Included at the lowest proficiency level at the grade of enrollment
Vermont	Yes (Score transformation rules)	AYP index- All assessment scores transformed to a 0-500 point scale	Included at the lowest proficiency level at the grade of enrollment

* Indicates that the state discontinued testing out-of-level in 2003-2004.

Table 11. Levels Testing Policies – Test Score Use by States

State	Equated to On-level Scores	Reporting Methods	Accountability Reporting
Kansas	Unknown	Unknown	Unknown
Oregon	No	Unknown	Unknown
Texas	No	Disaggregated	Unknown

Discussion

Just as was the case when we studied out-of-level testing policies in 2001 (Thurlow & Minnema, 2001), we found in this update that out-of-level or levels testing policies are rapidly changing. Since our 2001 study, we have made frequent updates to our policy files, and regularly checked state education agency's Web sites. By using data from the 2003-2004 school year (past implementation of out-of-level or levels testing), we hoped to circumvent any recent policy changes that would out date our reported information. We also offered states the opportunity to review the data included in this report prior to publication. Despite these safeguards and best efforts to gather precise and inclusive data, it is likely that some information is incomplete or inaccurate.

In comparing the data gathered in this study with the data gathered in the original study, we identified six points of discussion that focus on changes in policy or practice from the original

study. Even though the descriptive data from this study do not lead to conclusive statements, they do illuminate recent changes in out-of-level or levels testing, and highlight the probable future path of this assessment option.

States’ use of out-of-level or levels testing appears inconsistent with federal policy. NCLB requires assessing the maximum number of students with tests aligned to grade level achievement standards and allows out-of-level testing only as an alternate assessment aligned to alternate achievement standards if it meets the requirements for out-of-level testing set forth in federal regulations (Federal Register, 2003). Although nine states had discontinued testing out of level or using levels testing since the 2000-2001 school year, 13 states had either continued or introduced this testing practice to their assessment programs since 2000-2001. In fact, more states were using out-of-level or levels tests in 2003-2004 (17) than were in 2000-2001 (14) when the original study was conducted. Additionally, we could only study those states that indicated to us that they were testing out of level or using levels tests; there may have been more states using these testing practices that we were not aware of. It is of concern that the use of out-of-level or levels testing has increased in this decade despite federal legislation that severely limits this practice. Perhaps a “policy to practice” gap exists in that states’ may allow out-of-level or levels testing, but few districts actually implement this testing option due to the resulting difficulties in meeting federal guidelines. No matter what the reason for this increase in the use of out-of-level or levels testing, the increase is of concern and warrants further investigation.

States use a greater variety of out-of-level or levels testing classification terms. The original study discovered that states treated out-of-level testing differently within their large-scale assessment programs, classifying this testing option as a modification, accommodation, alternate assessment, or adapted assessment. In 2003-2004, nine different classification terms were used to label out-of-level or levels testing. Perhaps this increase in terms indicates a greater variety of out-of-level or levels testing use among the states. Or, perhaps it indicates a need to restructure out-of-level or levels testing to better meet federal guidelines. An increase in classification terms only complicates the process of locating out-of-level or levels testing information, and serves to confuse interstate conversations and proceedings regarding out-of-level or levels testing. More consistent classification terminology across states is needed to facilitate reliable comparisons between states’ policies and practices for both accountability and research purposes.

Some states have changed the qualifications for students allowed to use out-of-level or levels testing. Of the 17 states that continued their out-of-level testing policies throughout parts or all of the beginning of the decade, four states altered their qualification criteria for students allowed to be tested out of level. Two states (Arizona and Hawaii) placed greater restrictions on student qualifications, while two states (California and Iowa) placed fewer restrictions on student qualifications. Arizona no longer allowed students with 504 plans to be tested out of level, and specified that only students with IEPs who were labeled as significantly cognitively

delayed were allowed to use this testing option. Hawaii no longer allowed ESLL (English as a Second Language Learners, which is how English language learners are referred to in Hawaii) to participate in out-of-level testing, limiting this option to only students with IEPs. California extended its qualifications to include students with 504 plans, and Iowa did away with specific qualifications altogether to make the decision to test a student out of level a local school level decision. There was no consistent trend of increased or decreased allowances of students tested out of level, another indication of the unstable environment in which out-of-level or levels testing exists.

Some states have added more content areas assessed with out-of-level or levels tests. It appears as though there is a small trend for states to include science, social studies, and writing content areas in their out-of-level or levels tests if these content areas were not already part of their testing options. Three states (Connecticut, Louisiana, and Utah) added one or two of these content areas to their out-of-level tests since the first study was conducted. Additionally, some states (Kansas, Nebraska, and Tennessee) that began testing out of level or using levels tests since the first study have also included science and social studies in these tests. It benefits those students tested out of level or with levels tests to include in the out-of-level or levels assessment all the content areas that are assessed in the regular assessment.

There has been an increase in state-level reporting of out-of-level or levels testing scores although states have tended to aggregate these results with on-level scores. There was great variability across the 12 states in reporting out-of-level test scores in 2001 (Thurlow & Minnema, 2001). Overall, no state reported out-of-level test data in state data reports in a clearly identifiable format that depicted below grade level test participation and performance. By understanding states' unique treatment of out-of-level test scores, it is possible to find these scores in some states' data reports. Six states (Alaska, Connecticut, Delaware, Louisiana, South Carolina, Utah) did include out-of-level test results in public reports. For instance, one of these states (Connecticut) reported the participation data for out-of-level tests but not the performance data. Another state (Louisiana) used an off-the-shelf norm-referenced test to test students below grade level so that the resulting test scores could be equated to on-grade level data. The remaining states disaggregated out-of-level test results in one way or another—but again, these data were not clearly identified as out-of-level test results.

In our most recent policy review, only two states indicated that they did not report out-of-level or levels test results at the state level in 2003-2004. Increased state-level reporting likely may have resulted from recent federal and state mandates requiring improved reporting practices for all students. Yet, it seems that many states aggregated out-of-level or levels testing data with on-level data, a practice that inhibits, if not eliminates, the possibility of identifying valuable student subgroup assessment information. If out-of-level or levels test results are included with on-level test results, it is impossible to determine how students tested out of level or with a levels

test performed on that test. States should strive to consistently and clearly report out-of-level or levels test results disaggregated from other results to accurately determine student performance and foster school improvement.

The long term effects of using out-of-level or levels tests remain unknown despite an increased use of high-stakes assessments. States have increasingly opted to include some form of a graduation exit exam in their large-scale statewide assessments (Center on Education Policy, 2003). These exams exist in many forms, from a grade 10 assessment of the statewide testing program to a special graduation assessment separate from the grade level assessments. When passing this exam is necessary to receive a high school diploma, these exams are considered high-stakes assessments. States that allow out-of-level or levels testing need to consider how to address these graduation exams for students who had previously taken below-grade-level assessments. Are these students still expected to take (and pass) this exam in order to graduate? Is there an alternative to the graduation exit exam for students using out-of-level or levels testing? In light of an increasingly high-stakes assessment environment, it is imperative that educators, state assessment personnel, and educational researchers all investigate the long term effects on students tested out of level or with levels tests. Further investigations will assist educators in making informed decisions when choosing out-of-level or levels testing as an option for an individual student, and assist policymakers in making informed decisions about the future of out-of-level or levels testing.

Final Thought

One of the major findings of this study is that the issues that surround out-of-level testing remain as contentious—if not more so—than when we conducted our first policy review study in 2001. The future of this testing option will remain of interest to practitioners, policymakers, and researchers at all levels of the American educational system.

References

Center on Education Policy. (2003, August). *State high school exit exams: Put to the test*. Washington, DC: Author.

Federal Register. (2003, December 9). *Title I -- Improving the Academic Achievement of the Disadvantaged, Volume 68* (236). Retrieved December 9, 2003 from <http://www.ed.gov/legislation/FedRegister/finrule/2003-4/120903a.html>

Thompson, S., & Thurlow, M. (2003). *2003 State special education outcomes: Marching on*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Available at <http://education.umn.edu/NCEO/OnlinePubs/2003StateReport.htm/>

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. Available at <http://education.umn.edu/NCEO/OnlinePubs/OOLT4.html>