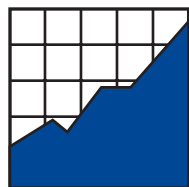




**Understanding Out-of-Level Testing in
Local Schools: A First Case Study of
Policy Implementation and Effects**



**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 11

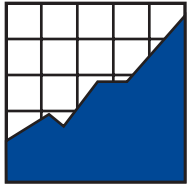
Understanding Out-of-Level Testing in Local Schools: A First Case Study of Policy Implementation and Effects

Jane E. Minnema • Martha L. Thurlow • Sandra Hopfengardner Warren

September 2004

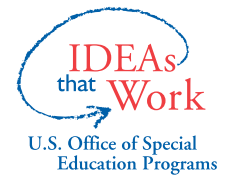
All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

Minnema, J. E., Thurlow, M. L., & Warren, S. H. (2004). *Understanding out-of-level testing in local schools: A first case study of policy implementation and effects* (Out-of-Level Testing Report 11). 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.

Overview

Standards-based instruction, with the aim of grade-level achievement for *all* students, is undoubtedly the most comprehensive educational reform of the recent past. A hallmark of this reform effort is the measurement of student academic achievement with large-scale assessments that are used for accountability purposes. Assessment results are to be made public as a way of accounting for the academic achievement of all subgroups of students. Just as teachers, parents, and students are interested in individual student achievement, policymakers and the public in general are interested in student group achievement that indicates how specific schools, school districts, and states are performing. Never before have schools and states been under such scrutiny for demonstrating improved student outcomes for specific subgroups of students – students with disabilities, English language learners, students receiving free and reduced lunch, and students in general education.

Today's emphasis on statewide testing that is used for accountability purposes has essentially been driven by federal mandates. The Elementary and Secondary Education Act (ESEA) of 1994 is a strong mandate that requires that all students with disabilities participate in states' standards-based assessment and be counted in states' accountability programs. Following a similar course in policy implementation, the Amendments to the Individuals with Disabilities Education Act (1997) first emphasized the inclusion of students with disabilities in large-scale assessment programs. Most recently, the re-authorization of ESEA, No Child Left Behind Act (NCLB) of 2001 has re-focused states' attention toward ensuring access to challenging, grade-level standards that are designed for students' grade of enrollment.

NCLB (2001) is currently the most stringent in requiring that all students be measured to grade-level criteria so that every subgroup of students receives challenging, standards-based instruction based on the grade in which these students are enrolled in school. Nevertheless, reviewing the chronology of federal law that has strengthened the inclusion of students with disabilities in states' large-scale assessment and accountability programs does not capture the political and controversial issues that have surrounded the implementation of federal mandates. This is certainly true for out-of-level testing, or the practice of testing students with disabilities below their grade of enrollment in states' large-scale assessment programs. Possibly no approach to testing has prompted such controversy at all levels of the American educational system—local, state, and federal—than out-of-level testing.

Out-of-Level Testing Background

Including all subgroups of students in statewide testing has been challenging for states. In order to administer more inclusive large-scale assessments, 14 states in 2001-2002 have added an

approach to their testing program known as “out-of-level testing” so that some students could be tested at test levels below their grade of enrollment (Minnema & Thurlow, 2003). Many arguments have been used to justify out-of-level testing. Policymakers, educators, and parents of students with disabilities thought that testing a student at the level on which they were instructed in the classroom would yield more accurate, precise, and useful test results (Thurlow, Minnema, Bielinski, & Guven, 2003; Minnema & Thurlow, 2003). It was also thought that testing students on their “instructional-level” would be less frustrating and embarrassing since students could fully engage in completing test items. Other commonly held beliefs about out-of-level testing included improved student motivation when taking tests, better attending behavior during test taking sessions, and enhanced student self esteem when students answered test items that tested content that they knew.

Also circulating in practice were attitudes and beliefs that discounted the value of out-of-level testing. While referencing different reasons, other policymakers, educators, and parents thought that out-of-level testing would not yield more accurate, precise, and useful test results because students were tested on test material that was developed for much younger students. Since the test material would most likely not be age appropriate, test motivation, attending behavior, and students’ self esteem could be adversely affected. Possibly the worst consequence of testing students with disabilities is the effects of setting lower expectations for students’ classroom performance or test level selection. In addition, public reporting of out-of-level test results was particularly problematic because data managers were unclear as to how to report the test scores—on the grade of the student’s test level or grade of enrollment in school.

The debate over the merit and worth of testing students with disabilities below their grade of enrollment continues to date. Researchers have begun to tease apart the complications of local and state level reporting, uneven policy implementation, the prevalence of below-grade level testing, and other such issues that surround the implementation of out-of-level testing policies results (Thurlow, et al., 2003; Minnema & Thurlow, 2003). Nevertheless, research has yet to weigh in on the factual basis of many of the beliefs, attitudes, and perceptions that surface in educational practice.

In order to understand how states actually administered out-of-level testing policies at the local level, we designed a case study to look closely at local educational agencies (LEAs) where students with disabilities were tested below their grade of enrollment. We also sought to determine whether the many popular beliefs in practice about out-of-level testing were actually true. To meet these aims, we implemented two research studies in two different school districts in two different states. Both of these states were administering out-of-level tests as part of their large-scale assessment programs during the school year 2001-2002 when we collected our data.

This report is the first accounting of one case study of large-scale assessment practices in a

local educational agency (LEA) where students with disabilities are administered states' standards-based tests out of level. A second report (Minnema, Thurlow, & Warren, 2004b) provides the write up of the results from the second case study conducted in another school district in another state. The overall purpose of our research project is to describe the specific effects of testing students with disabilities out of level as well as teachers' and students' perceptions of these effects.

State Context

In 2001-2002, the large-scale assessment program for the state chosen for the first case study was an augmented version of the Stanford Achievement Test, Ninth Edition (SAT-9) where items that directly measured state content standards were added to this norm-referenced test. More specifically, these additional test items were designed to measure students' progress on acquiring content standards in English-language arts, mathematics, science, and history/social science in grades 2 through 11. An augmented SAT-9 was included in the English-language arts and mathematics portions of the state test by selecting certain SAT-9 items that closely aligned with the state's content standards. The complete battery of the national norm-referenced SAT-9 was also given to students in grade 2 through 11, assessing reading, language (written expression), and mathematics. Students in grades 2 through 8 were assessed in spelling, and students in grades 9 through 11 were assessed in science and social science using the SAT-9. Also, state writing tests were administered in grades 4 and 7. In addition to the English version tests, the standards-based assessment and the SAT-9 tests, a Spanish Assessment of Basic Education, Second Edition (SABE/2) was used to assess Spanish-speaking students in reading, spelling, language, and mathematics in grades 2 through 11. These students must have been identified as a limited-English proficient student who had been in school for less than 12 months.

This state offered an out-of-level testing option for students with disabilities whose Individualized Education Program (IEP) documented a need for below grade level assessment. These students participated in the statewide testing program by taking the state standards-based test and the SAT-9 at any available level below the student's grade of enrollment. One level below the student's grade of enrollment was considered a standard test administration, while two or more levels below the student's grade of enrollment were considered a non-standard administration.

The School District

Data for this case study were collected in a unified school district located in the northern region of a large western state. The district served approximately 16,881 kindergarten through grade

12 students in 21 elementary, four middle, and four high schools. The mission of the district is to “produce educated citizens who achieve and perform at all levels of learning, are prepared to live fulfilling lives, and contribute to their community and the world in which they live.” The student ethnicity of the district includes 53% Caucasian, 38% Hispanic, 3% Filipino, 2% African-American, 2% Asian-American, and 2% American Indian.

Within this large school district, two middle schools (Schools 1 and 2) and one elementary school (School 3) were studied. School 1 served approximately 654 students in grades 6 through 8 who lived in a neighboring small city to that of School 2. Average class size was 27 students per class. School 3 was the most culturally diverse of the three schools studied, with a student population of 48% Caucasian, 19% Hispanic, 12% Filipino, 11% African-American, 3% Asian-American, 3% Pacific Islander, and 1% American Indian/Alaskan Native. Of the total student body, 8.7% of the students in School 3 were classified as limited English proficient. School 1 had four special education classrooms, two special day classes for students with cognitive disabilities and two classrooms for students with learning disabilities. Students from each of these classes were integrated into general education classes with levels of special education support as necessary.

School 2 was located in the city proper, and served approximately 1,013 students in grades 6 through 8. The average class was 28 students. There were more special education programs housed in School 2 than in the other two schools, two of which are a therapeutic day class for students with emotional and behavioral disabilities and a segregated classroom for students who have cognitive disabilities.

School 3, the elementary school, was located in a rural area of a relatively populated area of this state. Approximately 207 students attended this school in grades K-6. In School 3, the primary instructional focus was literacy for approximately 207 students in grades K-6. The curriculum was guided by this state’s content standards so that every child has the opportunity to learn grade level standards. Average class sizes for grades K-3 was 18 students while the average class size for grades 4-6 was 26 students.

Method

Research Questions

Our research project addressed the following research questions:

- (1) What are the instructional effects on students with disabilities who are tested out of level in statewide assessments?

- (2) What are teachers' learning expectations for students with disabilities who are tested out of level?
- (3) How are students with disabilities selected for an out-of-level test?
- (4) How do students with disabilities perceive out of level testing?

Sample

Our purposive sample included students with disabilities (n = 14), general education teachers (n = 5), special education teachers (n = 8), school administrators (n = 3), special education coordinators (n = 1), and other school staff such as guidance counselors (n = 1) and district test coordinators (n = 1). These participants were employed by a school district that was recommended by the state educational agency. The schools were self-selected by the district test coordinator and director of special education. Each participating school agreed to participate in the case study. All participants received a gift card for a local department store with the amount dependent upon the amount of time invested in our research activities.

Research Design

We used a case study design to address our research questions by employing mixed methods to garner numeric and narrative data.

Instruments

Our data collection techniques included face-to-face interviews (n = 33) and a document review of students' Individualized Education Programs (n = 14) and students' school records (n = 54). The data collection activity, face-to-face interviews, required approximately 25 minutes for school personnel to complete and less time for students with disabilities to complete. The purpose of these educator interviews was to garner their opinions about and their perceptions of student experiences in out-of-level testing while the purpose of the student interviews was to learn directly from them how they perceived their test experiences (see Appendix A for copies of the interview protocols).

Procedures

We designed our case study to collect interview and document review data on-site in each participating school and school district offices. One person from each school served as a contact person to assist in scheduling interview appointments and distributing the written surveys.

Data Analysis

We used two basic approaches to analyzing our case study data. For our qualitative data analysis, all educator interviews were tape recorded, transcribed, and subjected to a content analysis that yielded themes of results. Since the student interview responses were briefer than the educator interviews, these interviews were not tape recorded. Instead, student responses were written down during the interviews. To analyze the student interview data, we tabulated categories of responses rather than creating themes of results. In terms of our numeric data, we used descriptive statistics to analyze the IEP review data.

Findings

IEP Document Review

IEPs were reviewed in one middle school (School 1) where 25 students were tested out of level. Of these 25 students, 14 parents granted permission for us to review their child's IEP. Special education teachers provided some information for the remaining 11 students. In a second middle school (School 2), 54 students with disabilities were tested out of level. A special education coordinator provided some data for these students. In the remaining one school, only two students were tested out of level. These students were not included in the data collection activity because the small number does not support an analysis at the school level.

Table 1 shows the numbers of students tested out of level as a function of their grade, disability, and subject. Most students tested out of level in School 1 were assigned to the 6th grade; however, grade level was not provided for approximately half of the students. Within School 2 where we had grade assignments for all students, the majority of students tested out of level attended 7th grade.

In terms of disability category, most students tested out of level in both schools had learning disabilities. This finding may be due to the national trend in which students with learning disabilities is the largest category of students identified for special education services (U.S. Department of Education, 2001).

Regarding special education setting, more students attending Resource Classes in School 1 were tested out of level while more students placed in Special Day Classes were tested out of level in School 2. Some of the 7th grade students in School 2 attended core content classes in both resource and special day classes. Since these students are neither resource nor special day class students only, they are presented as "combined" in the table below. Combined placements such as these did not occur for the 8th grade students in School 2 or for any students in School 1. Several students from School 2 were placed in specialized therapeutic programs or alterna-

tive middle school classes. The number of students tested out of level from these programs is designated as “Other” in Table 1.

Table 2 demonstrates that the majority of out-of-level tests administered to middle school students were at least three grade levels below the grade in which the students were enrolled in school. In School 1, 19 of 25 tests (with 2 students missing a test level) were presented at the 2nd grade through 5th grade level. Only four students were tested at either 6th or 7th grade. While School 2 had 15 of 54 students with disabilities tested at the 6th grade level, 37 of 54 tested received state tests at either the 2nd, 4th, or 5th grade level.

Table 1. Number of Students Tested Out of Level by Grade, Disability, and Setting

	Grade 6	Grade 7	Grade 8	Missing Data	MR	LD	ED	Missing Data	Resource Class	Special Day Class	Combined	Other	Missing Data
School 1	11	8	3	3	10	14	0	1	14	10	0	0	1
School 2	2	33	19	--	2	40	12	0	10	24	8	12	0

Table 2. Grade Levels Administered as Out-of-Level Tests

Grade Level of Tests	School 1	School 2
2	4	13
3	3	--
4	6	20
5	6	4
6	2	15
7	2	2
Missing data	2	--

The data in Table 3 show differences between School 1 and School 2 in terms of the number of levels below grade level tested by out-of-level tests, and whether entire tests (or partial tests) were administered out of level. In School 1, all of the out-of-level tests were administered as entire tests. An equal number (n = 11) were presented 1 to 2 levels below grade level as were presented 3 to 5 levels below the students’ assigned grade levels. One test was administered 5 levels below grade level and no tests were administered 6 grade levels below. There were three students in School 1 with missing data. In School 2, 31 of 54 out-of-level tests were tested close to the students’ assigned grade level (i.e., 1 or 2 levels below). Of the 29 partial out-of-level tests, 16 were administered either 3 or 4 levels below grade level while 13 tests were administered either 5 or 6 levels below.

Table 3. Number Levels Tested Below Grade Level by Entire or Partial Test

Number Levels Below Grade Level	School 1	School 2	
	Entire Test	Entire Test	Partial Test
1	6	17	--
2	5	3	--
3	6	3	8
4	4	2	8
5	1	--	7
6	--	--	6
Missing data	3	--	--

Even though the number of IEPs reviewed (n = 14) is relatively small, the results remain interesting in that there is high variability between reading and math instructional grade levels when compared to the grade levels at which the students were tested. In this school, teachers indicated during the face-to-face interviews that out-of-level test levels were set according to students' academic strengths as demonstrated by academic progress in classroom performance. When comparing teacher determined instructional levels in students' reading and math skills to the grade levels at which they were tested in reading and math, most of the students (n = 13) were not tested at appropriate levels according to the grade level that teachers identified as their academic strength. For instance, one 6th grade student, whose teacher identified as reading at a 2nd grade level, was tested at the 3rd grade level even though math abilities were identified to be at the 5th grade level. When using the criterion of testing a student at his or her teacher-identified grade levels in reading and math, only one student's test levels matched his or her instructional grade level because the teacher-identified reading and math levels were set at the same grade level.

Table 4. Grade, Content Area, and Test by Levels for School 1

	Student Grade in School													
	6	6	6	6	6	6	6	6	6	6	7	7	7	7
Reading Level (Teacher Identified)	2	--	3	3	4	--	3	5	1	3	--	2	3	5
Math Level (Teacher Identified)	5	--	2	--	5	--	--	4	4	--	--	--	3	5
Test Level	3	2	--	4	4	2	3	5	5	5	--	4	5	5

Face-to-face Interviews

Student Interviews

We interviewed middle-school students with disabilities (n = 10) who attended special day classes or resource classes. Some students were included in general education instruction with paraprofessional support. Interview results included the following:

- Most students said that they liked taking the statewide test out of level (8 out of 10 students). Half of the students thought it was neither too hard nor too easy, although two students described the below grade-level test as “too babyish.” When asked about test rigor, seven students mentioned guessing at item responses, although six of these students indicated that they guessed only minimally and one student guessed frequently. Only one student indicated guessing at all throughout the out-of-level test. Two students did not mention guessing at test item responses.
- Only four students were able to appropriately describe an out-of-level test as being a test at a grade level below that which they were enrolled in school.
- Only one student knew that someone on the IEP team made the decision to test out of level. But even though this student attended the IEP team meeting, the student did not know if a parent or teacher made the decision.
- None of the students’ responses indicated that they understood how out-of-level testing could affect their future school experiences. It is interesting to note that six of the ten students plan to graduate from high school with three planning to receive a regular high school diploma. Of these six students, three planned to attend college. Of the remaining four students interviewed, two of them have set post-high school occupational goals.

Special Education Teacher and Administrator Interviews

The results of our face-to-face interviews with teachers and administrators are presented by themes of results. The thematic results are divided into three topical areas: (1) comparing out-of-level testing to on-grade level testing, (2) selecting students and test levels for out-of-level tests, and (3) interesting aspects of policy implementation. Included in the teacher interviews were the three special education coordinators who also had caseloads of students with disabilities for whom they provided services.

Comparing Out-of-level Testing to On-grade Level Testing

In Table 5 we present the results from the interview questions focused on the benefits and concerns of out-of-level testing. These narrative data revealed that there are varied opinions about out-of-level testing that do not fall into an orderly pattern. Both teachers and other school staff (e.g., principals, guidance counselor) identified benefits and concerns about testing students with disabilities out of level. In one case, the same idea, “negative impact on students’ self-esteem,” was identified as a concern for both out-of-level testing and on-grade level testing. Of importance in the teachers’ responses is the concern that an out-of-level test does not document achievement toward grade-level standards. Teachers suggested further that this is particularly true if students are continually tested out-of-level at the same grade level. Teachers also highlighted the concern that “sometimes they [students with disabilities] are tested in one area that maybe is below their ability, but that’s the way the tests are given. They’re all given at one grade level.” Another concern that was reflected in teachers’ responses was the lack of usable test results because only raw scores are provided for out-of-level tests given at more than one level below students’ grades of enrollment. These scores do not provide the normative information necessary to make instructional decisions. It is interesting to note that administrators identified concerns about out-of-level testing even though this question was not posed to them. These concerns parallel the concerns raised by the teachers.

The responses to the interview questions that compared student behavior during out-of-level testing to on-grade level testing fell into a clear pattern of results when the test environment was considered (see Table 6). Teachers identified inappropriate test behaviors during both out-of-level testing and on-grade level testing. In contrast, teachers identified appropriate test behavior during out-of-level testing, but not during on-grade level testing. Inappropriate test behaviors were said to occur during out-of-level testing only when multiple levels of the same test were presented within the same classroom. In other words, when students could compare their test level to the test level of other students, their behavior tended to be disruptive. Some teachers noted that during this testing situation, some students appeared to feel badly about having a test level lower than the other students who were testing in that classroom. While this interview question was not part of the administrators’ interview protocol, some administrators commented that they were not aware of student behavior during out-of-level or on-grade level testing.

Selecting Students and Test Levels for Out-of-Level Tests

One of our research questions pertained to the selection of students with disabilities for out-of-level testing. In order to answer this research question accurately, it is important to first consider the educational context in which these assessment decisions were made. Most students in School 1 and School 2 who attended resource classes had learning disabilities and were participating

Table 5. Comparing Out-of-level and On-grade Level Testing: Benefits and Concerns

	Out-of-level Testing	On-grade Level Testing
Benefits	<p>Teachers thought that:</p> <ul style="list-style-type: none"> - Test items answerable. - Better test motivation. - Practice taking tests. <p>Other school staff thought that:</p> <ul style="list-style-type: none"> - Large academic gains documented over time. 	<p>Teachers thought that:</p> <ul style="list-style-type: none"> - Better challenge for students included in general education.
Concerns	<p>Teachers thought that:</p> <ul style="list-style-type: none"> - No new test information provided. - May be inaccurate measure of ability. <p>Other school staff thought that:</p> <ul style="list-style-type: none"> - Not useful for instructional decisions. - Negative impact on self-esteem. - Logistics were difficult. 	<p>Other school staff thought that:</p> <ul style="list-style-type: none"> - Poor test motivation. - Negative impact on self-esteem. - Reduces instructional time.

Table 6. Comparing Out-of-level Testing to On-grade level Testing: Student Behavior

	Out-of-level Testing	On-grade Level Testing
Benefits	<p>Teachers thought that:</p> <ul style="list-style-type: none"> - Students attentive and on-task. - Calm and focused. - Worked hard. - Better attitude. 	<p>No appropriate or inappropriate behaviors identified.</p>

in mainstream education with paraprofessional support during instruction. Generally speaking, students who received special education services in special day classes had more severe disabilities so that little to no instruction occurred in general education classrooms. Within this context, it seemed that an underlying assumption was driving the decision to test students with disabilities out of level in both schools. Teachers who taught special day classes generally believed what was reflected in the following statement made by a special day classroom teacher: “Most of our students are below grade level, so we know that a grade level test would be really hard for them to do, or next to impossible. Most of them are three to four years behind grade level. To give them one grade level lower doesn’t really help that much.” Another participant indicated that, “even if they’re in a regular class, sometimes the level of work they’re getting is not 7th grade or 6th grade level, but more like 4th or 5th grade.”

In terms of selecting students for out-of-level testing, each educator interviewed indicated that

the decision to test a student out of level was discussed and decided during the student's IEP team meeting. It generally occurred near the end of the meeting when the IEP paperwork was completed. The team case manager writes in the IEP the decision to test out of level that requires a parent initial indicating agreement. Participants indicated varying levels of active discussion in making the decision to test below grade level. For instance, one participant commented that "the IEP team determines that, but honestly, it would boil down to a lot of input from the special education team. A lot of it has been my decision." On the other end of the continuum, an administrator suggested that "it happens at the students' IEP, where the parents are involved. If the student is so delayed, where he's working two grade levels behind, then the topic is really raised."

According to the themes from our interviews, four factors were considered in determining whether a student should be tested on grade level or out of level. First, teachers thought about students' "ability levels" based on "what their [academic] strengths are." Typically, this was based on their "functioning levels in the classroom and their work samples." Second, "educational assessment results for individual students" are considered that point to specific grade levels of ability. Third, parent considerations are also part of the decision-making process, meaning that "the parents actually make the decision after we [special education teachers] counsel them on ... what grade level the kid is reading at and what we think he would do well on." "Sometimes parents will want this kid tested at grade level ... because they think it will help motivate the kid." But, "most of the time the parent goes with what we suggest." One participant suggested that in making the decision to test out of level, "most parents are worried ... so they go by how frustrated their child is." Finally, "if you go just one grade level down, the test still counts as a standard presentation. Sometimes that factors in."

There is also a generally accepted process among both teachers and administrators to select a grade level at which to administer an out-of-level test. When thinking about students' level of academic functioning in determining the need to test below grade level, the grade at which to test is also considered. Teachers do this in two ways. First, "it's based on their ability level," which is determined "by the assessments, either standardized or non-standardized, that I do and by the discussion with the teachers about their functioning level in the classroom." A test level is then selected according to "... what grade level we feel they could take the test and still be a little challenged, but also be able to succeed in it."

Students new to the school district tended to be exceptions to the IEP decision-making process. In one case, a special education teacher reported calling a parent of a new student on the telephone to say, "The [statewide tests] are coming up. Your child just tested at this grade level, and I think it would be a good idea for us to let her keep taking tests at that same level. The parent said, 'Oh. OK.'" This student was the only student in resource classes whose selection for out-of-level testing was reported as based on a teacher recommendation with parental passive

acceptance. Teachers and administrators also indicated that students new to this school district entered with the decision to test out of level already made by the previous IEP team.

Aspects of Policy Implementation at the School Level

There are also interesting aspects to the manner in which this assessment policy is implemented in the schools. For both schools, two patterns emerged in the teachers' and administrators' responses that highlight differences in testing practice according to students' educational placements and students' grade levels. These considerations are presented in Tables 7 and 8.

Table 7. Differences by Student Placement

	Special Day Class Students	Resource Class Students
Teacher Role	Recommendation determined prior to IEP team meeting.	Suggestion ready for IEP team meeting.
Parent Role	Passive acceptance of teacher recommendation.	Discussion of teacher suggestion with parent choice as final decision.

Table 8. Differences by Student Grade of Enrollment

	7 th Grade	8 th Grade
Teacher Preference	Out-of-level testing	On-grade level testing

Regarding students' educational placement, one special day class teacher from School 2 commented that "you're going to have two different answers ..." depending on whether students attend special day classes or resources classes. "In special day class, they choose first by saying, 'What is their strength? It is language arts, reading, or math?' Then they say, 'What level are they at?' If they're at the 2nd grade, they will take the 2nd grade level test. Generally speaking in special day class, they take only a partial test. We're trying to make it as minimal as we can to get through." This difference was also apparent in School 1, where students received entire state tests. A special day class teacher commented, "During the IEP meeting ... the facilitator will look at the teachers for our recommendation. We usually go into those meetings after thinking that out. We'll make a recommendation that goes to the parents. The parent hears it out, and never has a parent disagreed with our recommendation." However, for resource class teachers, responses included, "The parent actually makes the decision." Or, "When we have an IEP meeting, we talk with the parents about the pros and cons of out-of-level testing. The IEP team as a whole has a sense of what they would recommend, and we will tell the parent that, but we always tell the parent that ultimately it's going to be their decision."

In School 1, our narrative results indicated a clear pattern in teacher preferences for the level at which their students were assessed. When a student is in 7th grade, teachers appeared more open to testing further below the grade of enrollment. However, administrators said that by the time students were attending 8th grade, the need to prepare and pass the High School Exit Exam drives the decision to test a student on-grade level. “There’s a difference in 7th and 8th grade. Our 8th grade teachers want them taking it on grade level and our 7th grade teachers generally want them to take it a grade level below. We want them [8th grade students] to be compared to what they really need to know at this grade level so that we know what to work on. We know they’re going to have to do the exit exams.”

When the special education teachers in School 1 created their interview schedule to participate in our data collection process, they began discussing the selection process for testing students with disabilities out of level. Through this discussion, they learned that differences existed between the four teachers about how to select students for out-of-level tests and how to determine the appropriate out-of-level test level. Among the four teachers, two taught special education classes in language arts and two taught special education classes in math. When selecting students for an out-of-level test, language arts teachers thought about the entire test in terms of how well a student could read the language arts and the math test. One language arts instructor commented, “it [out-of-level testing] is usually based on their reading level. If they can’t read the test, they’re not going to perform well on it anyway.” Accordingly, language arts teachers tended to select test levels, again, by considering how well students could read the entire test. However, “... if the kid is doing 4th or 5th grade math, we explain to them [parents] that they’d have to take a 7th grade test. But they’re really working on 5th grade level math and it may be too difficult for them.” Both math teachers in School 1 considered students’ math abilities only in deciding to test students out of level. One math teacher, however, provided read aloud accommodations for all students for whom she administered the test “to make sure that reading abilities didn’t interfere with their performance.”

Table 9. Differences by Content Area of Instruction

	Language Arts	Math
Teacher Thinking	Considered reading level for entire test.	Considered grade level of student abilities in one content area only.

Discussion

The final step in the interpretation of our case study data was to seek out points of commonality and corroboration between our interview data from multiple sources and our review of students’ IEPs. We discuss these points through “grand themes” that emerged from our analysis as over-

arching findings that we believe to be important considerations for policymakers and educators. If alternative view points emerged in our data sets, those findings are presented as caveats to our grand themes of considerations.

Students with disabilities who were tested out of level were not instructed on the grade level in which they were enrolled in school.

As a first consideration, our findings suggest that all of the students tested out of level in these two schools were not receiving standards-based instruction in all content areas that was commensurate with their grade of enrollment. Our narrative results support this conclusion in that teachers believed that some students in special education will never be able to meet grade level standards. These comments were taken from interviews with teachers who taught students whose disabilities could be considered mild or moderate since each of these students receive general education instruction for a part of each school day. In addition, both teacher-provided information about students' reading and math achievement levels and IEP review results indicated that some middle school students were achieving at elementary grade levels. This conclusion points to two issues that are important considerations for policymakers and practitioners.

First, since NCLB mandates that *all* students are to receive grade-level, standards-based instruction, it is important for educators to think carefully about how to bring all students up to proficient levels of performance. This issue is particularly critical for students with disabilities who have not been receiving grade-level instruction in the past and may have to acquire more content in less school time than would normally be expected. Second, since test results were not available at the time of our document review, we do not know how these students performed on the state test administered below grades of enrollment. If the majority of these students either passed or nearly passed the statewide test, it is possible that some of these students may have received standards-based measures that did not adequately challenge their actual academic abilities.

Of the students tested out of level in these two schools, only 30% of the test scores were entered into accountability indexes.

A second issue that arises from our final interpretations concerns out-of-level test score use for system accountability purposes. State policy defines one grade level below a student's grade of enrollment as the only out-of-level test presentation that can be considered to be a standard test administration. This means that those test data from statewide assessments administered more than one grade level below were *not* used in calculating school systems' academic progress from year to year. Given this policy constraint, only 23 of 76 out-of-level tests overall from Schools 1 and 2 could be used for public reporting of test performance. Those students with disabilities ($n = 53$) who were tested more than one grade level below their grade of enrollment were not included in the system level calculation when accounting for schools' academic achievement progress to the state and the public. Also, school system planners were not able to consider these students' academic needs since their test performance was eliminated from school system

“Alternative Finding 1”

A teacher new to the profession provided an alternative perspective on the learning capabilities of students with disabilities. She taught a segregated special education class, referred to as a special day class, containing students with severe disabilities. During her interview, she commented that she “viewed her job” to be one of “getting my students [with severe disabilities] into the resource program.” Students progressed out of her segregated classroom into a resource classroom where students could be included in general education instruction. In fact, one of her students whose reading disability is severe, was included in general education math classroom instruction because only reading skills lagged behind those skills of his grade level peers. Both instruction and classroom tests are accommodated, which is consistent with the student’s IEP. Communication between the special education and general education math teachers is unique since they are married. It was also noted that the “student works really hard” and that his parents “are very supportive.” Nonetheless, the student with a severe disability is acquiring math content on the grade in which he is enrolled in school due to teacher, parent, and student high expectations for his learning!

calculations. Now that schools need to demonstrate adequate yearly progress (AYP)—with planful responses for improving student achievement—policymakers and educators alike will be hard pressed to do so when specific subgroups of students are not included in system accountability programs.

Some out-of-level test results are not presented in a usable test score form.

Our case study results point to a third overarching concern, one that relates to test data use and interpretation. The test contractor that provides the state test does not analyze out-of-level test scores that are more than one grade level below a student’s grade of enrollment. In other words, teachers and parents receive as a student test performance report a raw number that provides no comparative numeric information. The raw number represents how many items were answered correctly, but provides no additional information as to which items were correct or how this test score compares to test scores of other students. In other words, there are no contextual analysis features accompanying an out-of-level test report with which to interpret a given student’s performance. There was one special education teacher in particular whose responses supported our grand theme. She was “frustrated because the test scores have no normative information so that we can’t compare our students to other students in our school district or the state.”

Of concern to policymakers and educators is the manner in which the test contractor prepares the results of out-of-level tests. The test score analysis procedures in this state manage one subgroup of the student population, namely students with disabilities who are tested more than one grade level below the grade in which they are enrolled in school, differently than the remaining

“Alternative Finding 2”

Through probing interview questions with a few of the special education teachers from School 1, we learned more specifically how students with disabilities were selected for out-of-level testing. Within some of these teacher responses, we learned that some teachers factored into their decision making whether the test score would be useful for accountability purposes. While many of the out-of-level test scores were not reportable test data, it was encouraging to learn first that teachers understood the consequences of administering a nonstandard version of the statewide test and second, that this was a consideration in selecting a student for below grade-level testing.

student population. When this occurs, practitioners, school planners, parents, and the students themselves will not be able to accurately follow individual and group academic progress toward acquiring grade-level content standards. Understanding individual student progress is a necessary component of improving instructional delivery, so that educators can ensure maximum benefit from standards-based educational reform for all students.

Uneven out-of-level testing policy implementation occurred within and between schools.

Our results highlighted inconsistencies across teachers within one school and between the two schools as they implemented the state’s out-of-level testing policy in their schools. In terms of within school differences, the decision-making process used to select test levels for out-of-level state tests was approached differently by different special education teachers. For School 1, special education teachers who taught different curricular content selected test levels according to students’ performance in the content area for which they were responsible. Since only one special education teacher attended a student’s IEP team meeting and did not consult with other teachers prior to the meeting, an out-of-level test level was based on the achievement in one content area only. Differences between schools also emerged in our data sets. Our document review indicated that in School 2 some students who were tested below grade level took partial standards-based assessments while students in School 1 always took the complete test albeit below their grade of enrollment.

Both of these examples from our participating schools point to policy implementation inconsistencies neither of which matches the intent of the state-level policy content. Uneven policy implementation across geographic regions is difficult to avoid, particularly when the area is expansive. Under these conditions, ensuring appropriate state-level policy implementation is generally challenging for all involved. Policymakers are prompted to continually strive to provide multiple formats of the most comprehensive and up-to-date training programs that reach as many practitioners as possible.

“Alternative Finding #3”

Teacher responses to our interview questions did not always corroborate our interpretation of the document review findings. Most special education teachers indicated that they were able “to use out-of-level test scores by comparing a student’s test score to the test scores from previous years.” In doing so, they were able to see if learning was progressing from one year to the next. We interpreted these contradictory opinions as indicative of teachers’ varying abilities to interpret and apply test data—rather than findings that contradict our grand theme. None of the special educators who indicated that they used raw scores for instructional purposes commented about test normative data that accompany the state’s standards-based assessment. While these narrative data do not necessarily support our grand theme, our opinion concerning the lack of usefulness of raw test scores for accountability purposes is not in question.

Some students with disabilities who are tested out of level appear to be experiencing lost opportunities to learn.

Finally, the intent in administering an out-of-level test is to provide an appropriate large-scale assessment experience for all students with disabilities. Our case study results demonstrated that unintended consequences occur where students with disabilities are not always provided with test levels that tap academic abilities accurately. Some students were assessed further below grade level than they were instructed in either one content area or both content areas tested. In other cases, a pattern emerged in our data where students were tested further and further below grade level as they grew older. Students with disabilities who attended middle school were tested at early elementary grade levels. It is impossible to know how these students could have achieved without being provided opportunities to acquire grade level content standards through the delivery of challenging instruction that supports academic proficiency.

Of further concern is a major finding from our face-to-face interviews with students with disabilities. Passing a mandatory high stakes exam is required to receive a high school diploma, which is undoubtedly not possible when students are achieving at elementary grade levels. Our student interview results indicated that even though the majority of these students plan to graduate from high school, none of them understood that out-of-level testing does not promote grade-level standard achievement. When this information was shared with two special education teachers, a conversation regarding “students’ unrealistic goals for the future” ensued. Not only are lost opportunities to learn highlighted by these findings, but also teachers’ expectations for the levels at which students with disabilities can achieve academically. It behooves policymakers to think critically about unintentional consequences of policy decisions that play out in practice in ways that counter the purpose of policy content.

“Alternative Finding #4”

Teachers in both Schools 1 and 2 appeared to carry assumptions about the learning potential for students with disabilities. One teacher commented that the reason that “out-of-level testing is a good idea is because special education students will never learn like other students.” During other interviews, special education teachers appeared reluctant to pursue how students with disabilities might be able to learn grade level curriculum. While our overarching interpretation of these data suggests that students who are receiving special education services in these schools are probably losing opportunities to learn at levels more commensurate with the same-age peers, some of the special education teachers we interviewed assumed that this possibility was unlikely due to their preconceived understandings of the limited learning potential for those students identified with disabilities.

Conclusion

Since these data were collected, the state mandated out-of-level testing policy has undergone major revisions. Undoubtedly due to federal discouragement of using out-of-level testing in lieu of states’ regular or alternate assessments, the state educational agency decided to first limit, and then phase out the use of below grade-level testing. In the first year, out-of-level test levels were limited to only one grade level below a student’s grade of enrollment. Next, over the three school years, out-of-level testing was to be eliminated from large-scale assessment practices. In light of the current federal mandate requiring that *all* students receive challenging, grade-level, standards-based instruction, our findings are particularly useful for this state as well as other states who are striving to meet the mandates of NCLB.

Given our case study research design, our results cannot be generalized to other school districts within the state that data were collected. Yet, our findings do point to key concerns that can be raised with educators when working with schools throughout this state as well as other states. Even though this case study focuses on a limited number of participants that are purposely selected, our findings accentuate the need for policymakers, educators, and parents to think critically about the immediate and long term unintended consequences of testing students with disabilities out of level in states’ large-scale assessment programs. These concerns are especially relevant as states strive to demonstrate increased student proficiency as measured by standards-based measures that are administered at the grade level in which students are enrolled in school.

References

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. Available at <http://education.umn.edu/NCEO/OnlinePubs/OOLT10.html>

Minnema, J., Thurlow, M., & Warren, S. (2004b). Understanding out-of-level testing in local schools: A second case study of policy implementation and effects (Out-of-Level Testing Report 12). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Available at <http://education.umn.edu/NCEO/OnlinePubs/OOLT12.html>

Thurlow, M., Minnema, J., Bielinski, J., & Guven, K. (2003). *Testing students with disabilities out of level: State prevalence and performance results.* (Out-of-Level Testing Report 9). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Available at <http://education.umn.edu/NCEO/OnlinePubs/OOLT9.html>

Appendix A

Instruments

Interview Protocols

Teacher

Principal

District Test Coordinator

Student

Teacher Face-to-Face Interview Protocol

“I am _____ from the University of Minnesota. Your school district has agreed to participate in one of our research studies that is collecting data to understand the effects of testing students with disabilities out of level in large-scale assessments. Part of that research study is our interview. Like we discussed before, I have seven questions to ask you about out-of-level testing in large-scale assessments. I’d like to tape record our conversation if that is all right with you. That way, I will have exactly what you have said to make sure that I don’t make any mistakes when I analyze the responses to these questions. Before we begin however, I have a consent form that I would like for you to read and then sign if agreeable.”

“Thank you. Do you have any questions before we begin?”

Q1) Do you think that out-of-level testing is beneficial for your students? If so, why? If not, why not?

Q2) Do you think that on-grade level is beneficial for your students? If so, why? If not, why not?

Q3) How did your students with disabilities behave when taking an on-grade level test? How did they behave when taking an out-of-level test?

PROBE: Did any of your students act out when taking a test on-grade level?

Q4) How do you think students felt about taking a test out of level? How do you know this? Did your students comment about the test booklet? Did the test material seem age appropriate?

PROBE: Did your child think that the out-of-level test was appropriate for his/her age? If so, how do you know?

Q5) Who actually decides which students with disabilities take an out-of-level test?

Q6) Can you please describe how the decision to test a student out of level is made?

Q7) How do IEP teams determine the appropriate level of an out-of-level test? Does the test level typically align with a student's instructional grade level? Are test levels ever assigned by the level at which a student is certain to be successful? Can teachers identify the grade level of a test by looking at the content of the test items?

Q8) Do any of your school staff, including the administrators, advise you about out-of-level testing? If so, what do they say?

Q9) How will taking the state test out of level affect your student(s) in the future?

PROBE: Is something being done to make sure that your students are catching up to grade level standards?

Q10) Do you think that the student's parent(s) understand the consequences of taking the state test out of level? Do you think that the student who is tested out of level understands the future consequences of taking the state test out of level?

Q11) Are you familiar with the public reporting of state test scores in your community? I have a question that asks for your opinion from three choices. I assume your students' names are kept confidential. When test scores are reported to you, the family, and the public, would you like the out-of-level test scores to be compared to:

(√ one)

- ___ The grade level of the out-of-level test?
- ___ With the grade level of his/her classmates?
- ___ No opinion.

Please explain why?

Q12) How do you interpret an out-of-level test score? How do you use out-of-level test scores? Is there a difference in how you use out-of-level test scores and in-level test scores?

Principal Face-to-Face Interview Protocol

"I am _____ from the University of Minnesota. Your school district has agreed to participate in one of our research studies that is collecting data to understand the effects of testing students with disabilities out of level in large-scale assessments. Part of that research study is our interview. Like we discussed before, I have seven questions to ask you about out-of-level testing in large-scale assessments. I'd like to tape record our conversation if that is all right with you. That way, I will have exactly what you have said to make sure that I don't make any mistakes when I analyze the responses to these questions. Before we begin however, I have a

consent form that I would like for you to read and then sign if agreeable.”

“Thank you. Do you have any questions before we begin?”

Q1) Do you think that out-of-level testing is beneficial for your students with disabilities? If so, why? If not, why not?

PROBE: Do you know if any students acted out when taking an on-grade level test?

Q2) Who actually decides which students with disabilities take an out-of-level test? Can you please describe how the decision to test a student out of level is made?

Q3) Do you think that your IEP teams consider the future consequences of testing students with disabilities out of level? If so, how do you know? Do you think that the parents understand the consequences of testing students with disabilities out of level? Do the students who are tested out of level?

Q4) Do you, or anyone else, advise your teachers about out-of-level testing? If so, what kinds of things are said? How is the information prioritized? Does anyone advise you about out-of-level testing?

Q5) How do IEP teams determine the appropriate level of an out-of-level test? Does the test level typically align with a student’s instructional grade level? Are test levels ever assigned according to a student’s level of success?

Q6) Can you please describe what happens to out-of-level test scores after a student has completed the test? How are these scores included in school reports? In district reports? In state reports? How are out-of-level test scores used in school improvement plans? How do students benefit from school improvement plans?

Q7) How are out-of-level test scores used by your staff? Are these scores used for student accountability purposes? For system accountability purposes?

Q8) I have a question that asks for your opinion from three choices. In asking this question, I assume students’ names are kept confidential. When test scores are reported to you and to the public, would you like for your student’s test scores to be compared to:

(√ one)

- ____ The grade level of the out-of-level test?
- ____ The grade level of his/her classmates?
- ____ No opinion.

Please explain why?

Special Education Coordinator Face-to-Face Interview Protocol

“I am _____ from the University of Minnesota. Your school district has agreed to participate in one of our research studies that is collecting data to understand the effects of testing students with disabilities out of level in large-scale assessments. Part of that research study is our interview. Like we discussed before, I have seven questions to ask you about out-of-level testing in large-scale assessments. I’d like to tape record our conversation if that is all right with you. That way, I will have exactly what you have said to make sure that I don’t make any mistakes when I analyze the responses to these questions. Before we begin however, I have a consent form that I would like for you to read and then sign if agreeable.”

“Thank you. Do you have any questions before we begin?”

Q1) Do you think that out-of-level testing is beneficial for your students with disabilities? If so, why? If not, why not?

PROBE: Do you know if any students acted out when taking an on-grade level test?

Q2) Who actually decides which students with disabilities take an out-of-level test? Can you please describe how the decision to test a student out of level is made?

Q3) Do you think that your IEP teams consider the future consequences of testing students with disabilities out of level? If so, how do you know? Do you think that the parents understand the consequences of testing students with disabilities out of level? Do the students who are tested out of level?

Q4) Do you, or anyone else, advise your teachers about out-of-level testing? If so, what kinds of things are said? How is the information prioritized? Does anyone advise you about out-of-level testing?

Q5) How do IEP teams determine the appropriate level of an out-of-level test? Does the test level typically align with a student’s instructional grade level? Are test levels ever assigned according to a student’s level of success?

Q6) Can you please describe what happens to out-of-level test scores after a student has completed the test? How are these scores included in school reports? In district reports? In state reports? How are out-of-level test scores used in school improvement plans? How do students benefit from school improvement plans?

Q7) How are out-of-level test scores used by your staff? Are these scores used for student accountability purposes? For system accountability purposes?

Q8) I have a question that asks for your opinion from three choices. In asking this question, I assume students’ names are kept confidential. When test scores are reported to you and to the public, would you like for your student’s test scores to be compared to:

(√ one)

- ____ The grade level of the out-of-level test?
- ____ The grade level of his/her classmates?
- ____ No opinion.

Please explain why?

District Test Coordinator Face-to-Face Interview Protocol

“I am ____ from the University of Minnesota. Your school district has agreed to participate in one of our research studies that is collecting data to understand the effects of testing students with disabilities out of level in large-scale assessments. Part of that research study is our interview. Like we discussed before, I have seven questions to ask you about out-of-level testing in large-scale assessments. I’d like to tape record our conversation if that is all right with you. That way, I will have exactly what you have said to make sure that I don’t make any mistakes when I analyze the responses to these questions. Before we begin however, I have a consent form that I would like for you to read and then sign if agreeable.”

“Thank you. Do you have any questions before we begin?”

Q1) Do you think that out-of-level testing is beneficial for your students with disabilities? If so, why? If not, why not?

Q2) Do you, or anyone else, advise your teachers about out-of-level testing? If so, what kinds of things are said? Does anyone advise you about out-of-level testing? If so, what kinds of things are said?

Q3) Who actually decides which students with disabilities take an out-of-level test?

Q4) Can you please describe how the decision to test a student out of level is made? What are the steps that you go through so that a student can take an out-of-level test?

Q5) How do IEP teams determine the appropriate level of an out-of-level test? Does the test level typically align with a student’s instructional grade level? Are test levels ever assigned according to a level at which a student is certain succeed?

Q6) Can you please describe what happens to out-of-level test scores after a student has completed the test? How are these scores included in school reports? In district reports? In state reports?

Q7) How do you interpret an out-of-level test score? How are out-of-level test scores

used by your staff? Are these scores used for student accountability purposes? For system accountability purposes?

Student Face-to-Face Interview Protocol

“Hi. My name is _____ and I am from Minnesota. I have been in your school this week to learn more about out-of-level testing. Do you know what that is? Good.”

If not, continue with ... *“Do you remember when you took the (name of test) with all of the other students in your school? Do you know if your test was the same test as other (8th or 10th graders)? Good.”*

“Do you mind if I ask you a few questions about that test? The questions are easy. I’m sure that you will do very well. It’s not a test! It’s for a research study that I am doing. When we are finished I have a gift card for you to spend at Target. First, I need for you to listen to me read this paper. Then, if you want to answer my questions, I will need for you to sign this paper.”

“Do you have any questions before we begin?”

Q1) Do you like the (test name)? Why or why not? Did it seem okay for your age?

Q2) Do you know what out-of-level testing is? If a friend asked you what an out-of-level test is, what would you say?

Q3) Do you know who decided that you should take the test (use student’s words to describe test)? Did you help make that decision?

Q4) Do you know how taking this test (use student’s language) will change anything for you in school when you are older?”

“You’ve done a very good job answering my questions. That’s great! Enjoy spending your gift card at Target. Have a good rest of the day. Thank you.”