

# A Baseline Perspective on Disability Subgroup Reporting

Martha L. Thurlow and Hilda Ives Wiley

*National Center on Educational Outcomes, University of Minnesota*

Federal legislation requires states to publicly report on the participation and performance of students with disabilities on statewide assessments. Publicly reporting this information means that it will be more accessible to educators and laypeople, and it can be used in evaluating educational programs for and assessing the progress of this subgroup. In this study, the authors collected disability subgroup assessment information reported by the 50 states for both the 2000–2001 and 2001–2002 school years, essentially establishing a baseline for No Child Left Behind disability subgroup reporting. They examined the information that was publicly reported for general and alternate assessments. In 2001–2002, 48 states reported either participation or performance information for students with disabilities on at least one state assessment. However, only 35 states reported both participation and performance for *all* of their assessments, as required by the Individuals with Disabilities Education Act.

Current education legislation stresses the importance of including students with disabilities in state accountability systems and disaggregating their performance results. This marks an important shift from past practices of reporting only information about how all students performed on statewide assessments and, in some cases, of specifically excluding the performance of certain subgroups. Disaggregation allows the public to not only see how *all* students are performing but also assess which groups are making adequate progress.

Public reporting of student assessment data serves several purposes (Cibulka & Derlin, 1995). One purpose is to use the results to make informed decisions about educational programs and school effectiveness. Another purpose is to evaluate how well students are achieving. An effective accountability system should share results with both educators and the general public. Data should be presented in a clear, accessible format so that all stakeholders can determine which programs are effective.

Until the mid-1990s, large percentages of students with disabilities were excluded from large-scale assessments (Koretz & Hamilton, 2001; McGrew, Thurlow, & Spiegel, 1993). In 1994, however, the Elementary and Secondary Education Act (ESEA) amended Title I and required states to (a) assess students with disabilities using the same tests as were used with students without disabilities and (b) disaggregate scores for students with disabilities. Title I also mandated that results for students with disabilities and all students be reported in a public report of school progress (U.S. Department of Education, 1999). In addition to ESEA 1994, the 1990 Individuals with Disabilities Education Act (IDEA) was amended in 1997 to require that each state report assessment information for students with disabilities with the same frequency and in the same detail as for students without disabilities. This information is to

be reported for both general assessments and alternate assessments. States were required to develop alternate assessments for students unable to participate in general assessments, which most states interpreted to mean students with significant cognitive disabilities. States must report the number of children participating in general and alternate assessments, as well as the performance of those students on the assessments (National Research Council, 1999).

The 2001 federal reauthorization of the Elementary and Secondary Education Act, the No Child Left Behind (NCLB) Act, required that by the beginning of the 2003–2004 school year, assessment reports be completed at the classroom, district, and state levels (Fast, Blank, Potts, & Williams, 2002). Both states and districts must publicly report assessment results before the start of each school year. Because NCLB also holds schools accountable for specific groups of students, as well as the total student population (Koenig, 2002), states will likely increase their focus on students with disabilities and ensure that a high percentage of these students are participating in the assessment process.

Despite the importance of public reporting, there is a surprising lack of information and research in this field (Thompson, Thurlow, & Lazarus, 2001). In their search, Thompson et al. found that some literature addressed school improvement and state data requirements in relation to reporting. Most of this literature was directed at school improvement. Although the literature frequently described how states can use performance reports, it rarely focused on what information should be reported. This lack of information leads to confusion among states about how to present and format assessment results.

In a report addressing the six essential principles of inclusive assessment and accountability systems, Thurlow, Quen-

emoen, Thompson, and Lehr (2001) focused on reporting as one of the principles: "Principle 3. All students with disabilities are included when student scores are publicly reported, in the same frequency and format as all other students, whether they participate with or without accommodations, or in an alternate assessment" (p. 3). This principle is further delineated through four characteristics that clarify how all students are reflected in reports, even those students who might not achieve scores that can be calculated.

Educational reform policies stress the importance of including all students in accountability systems. Federal requirements such as IDEA and NCLB were created partly in response to findings of nonparticipation in assessments. For example, in 1992, only 19 states reported assessment information, with most reporting either participation rates of less than 10% for students with disabilities or that they did not know how many students with disabilities participated (Shriner & Thurlow, 1992). It is not possible to obtain an accurate picture of student performance if 10% or more of the population is not being assessed or if scores are not reported (Thurlow, Elliott, & Ysseldyke, 1998; Thurlow, House, Boys, Scott, & Ysseldyke, 2000; Thurlow & Thompson, 1999; Zlatos, 1994). Also, excluding students with disabilities from state assessments deprives these students and their parents from gaining knowledge about academic growth and progress toward state standards (National Research Council, 1999).

In general, in the early 1990s, state guidelines that addressed the participation of students with disabilities in statewide assessments varied across states (Thurlow, Ysseldyke, & Silverstein, 1995). In addition, the estimated participation rate of students with disabilities varied from one state to another, and rates were quite low (Erickson, Thurlow, & Thor, 1995; McGrew, Thurlow, Shriner, & Spiegel, 1992; Shriner & Thurlow, 1992). Early researchers concluded that to obtain a complete picture of how all students are performing, participation information and performance results for *all* students must be made available for public scrutiny. They concluded that if laypeople and educators could see that students with disabilities were not gaining skills or were not even being assessed, questions would begin to be asked about what could be done to ensure that all students are learning.

Over the past several years, an increasing number of states have begun to publicly report information about the participation and performance of students with disabilities in their statewide assessment programs (Bielinski, Thurlow, Callender, & Bolt, 2001; Thurlow, Langenfeld, Nelson, Shin, & Coleman, 1998; Thurlow, Wiley, & Bielinski, 2003; Ysseldyke et al., 1998). After the 1997 IDEA reauthorization, which required all states to publicly report participation and performance information for students with disabilities, states slowly began providing this information. However, not all of the states reported the results in a format that is clearly accessible to the public. For example, when Thurlow, Wiley, and Bielinski (2003) accessed data from 2000–2001, only 55% of Web-based assessment results had information about students with disabili-

ties posted with information for all students or students in general education. For the 2001–2002 school year, this percentage rose to 69% (Thurlow & Wiley, 2004). Furthermore, on their home pages, many states used phrases such as "Publications and Newsletters" and "Divisions and Units" that did not clearly indicate a link to disaggregated test results. With an increase in online reporting, states must be able to present assessment information that is easy to locate and comprehensible to laypeople and educators.

In this study, we sought to identify the extent to which states reported student participation and performance information for general and alternate assessments for the testing years 2000–2001 and 2001–2002, which essentially would serve as a baseline perspective on disability subgroup reporting. Specific examinations included what types of tests were being administered, how states reported participation information, and which states were publicly reporting both the participation and performance of students with disabilities.

## Method

For both years, we reviewed state Department of Education Web sites during the 2001–2002 and 2002–2003 school years to determine what assessment information was publicly available for the previous testing year. Participation and performance data for students with disabilities and all students were collected for all statewide K–12 assessments. Once we had collected the data from the Web sites, we mailed a letter to each state assessment director that outlined our findings for his or her state based on data posted on the Web site. Included in this letter was a table that outlined the names of the state's assessment(s), the grades and content areas tested, and the availability of disaggregated participation and performance results for students with disabilities. We requested that any additional information or changes to the information be sent to us. Because we were collecting data in the form in which they were publicly reported, many states mailed us printed reports of the assessment results. Other states provided dates by which the results would be posted online. We then e-mailed or called states for which we still had not found or been sent disaggregated results for any of their assessments. For 2000–2001 data, May 31, 2002, was set as the deadline after which we would no longer accept information. The deadline for 2001–2002 data was June 27, 2003. These dates gave states a full year to compile their assessment results.

To ensure that our findings were as accurate and complete as possible, we sent a similar letter to each state's director of special education. Letters were mailed to the states for which we had at least one missing piece of disaggregated data. In 2000–2001, we mailed letters to 23 state directors of special education (10 responded), and in 2001–2002, we mailed letters to 33 state directors (25 responded). These letters asked the directors to review the information and to alert us to any incorrect information by September 20, 2002/August 29, 2003,

respectively. In the 2 months after the deadlines, telephone calls were placed to follow up with those states for which we had not received a response from either the director of assessment or the director of special education.

Data collection and verification for alternate assessments were additions to the study during the first study year. We mailed a letter in May 2002 to every director of assessment and asked whether his or her state had reported data on an alternate assessment administered to students during the 2000–2001 school year. We received 28 responses. We then examined the state Web sites of those states from which we did not receive a response. If we found no publicly reported participation and performance results, we assumed that they did not report this information. For the 2001–2002 school year, we looked for alternate assessments in our Web search and included alternate assessment questions in the letters to directors.

For our analysis of states' reporting, we examined all tests available. For summary purposes here, we pulled data from just one grade for each school level (elementary, middle, high school) in each state. Whenever possible, we used fourth grade to represent the elementary level, eighth grade to represent the middle school level, and 10th grade to represent the high school level. We chose these grades because they are the grades at which the largest numbers of states test students. If a state did not assess children at those grades, we selected the grade below that level (e.g., third grade was selected if fourth-grade data were not available). If data for neither of those grades were available, we used the other remaining grade (e.g., fifth grade for elementary school if there were no fourth-grade or third-grade data).

## Results

After analyzing the types of tests that states administered, we identified 118 separate statewide tests or testing systems for 2000–2001 and 111 for 2001–2002. For 2001–2002, 35 states had more than one assessment, and every state had a state-mandated assessment (although Iowa's assessment system allows voluntary participation). Table 1 breaks down the testing systems by type: norm-referenced tests (NRT), criterion-referenced tests (CRT), exit tests used as a gate for graduation or earning a particular type of diploma (EXIT), and tests that combined standardized NRTs with additional state-developed test items (NRT/CRT). Although we recognized that exit exams may also be NRTs, CRTs, or both, the high-stakes consequences for students on these exit exams indicated a need to examine these tests separately. There was little change in types of tests given from 2000–2001 to 2001–2002. Just over half of all assessments were criterion-referenced tests, and the next most common (half as many) were norm-referenced assessments. Approximately one fifth of all tests were high-stakes exit exams. Only 4% to 5% were assessments that combined both criterion-referenced and norm-referenced components.

We examined the amount of information states provided publicly about their assessment participation and performance

**TABLE 1.** Types of General Assessments in 2000–2001 and 2001–2002

Type of test	2000–2001 (%)	2001–2002 (%)
Criterion-referenced test (CRT)	51	52
Norm-referenced test (NRT)	23	22
Exit (or graduation) test	22	21
NRT/CRT (combination)	4	5
Total	118	111

**TABLE 2.** Numbers of States That Reported Disaggregated Data for Students With Disabilities in 2000–2001 and 2001–2002

Nature of reporting	2000–2001	2001–2002
<b>General assessment</b>		
Performance and participation data for all tests	28	35
Performance data only for all tests	4	4
Performance and participation data for some tests	10	9
Performance data only for some tests	2	0
No disaggregated data	6	2
<b>Alternate assessment</b>		
Performance and participation data	13	22
Performance data only	2	5
Participation data only	10	5
No disaggregated data	25	18

results. We then grouped states into five categories according to what data they publicly reported:

1. performance and participation data for all tests,
2. performance data only for all tests,
3. performance and participation data for some tests,
4. performance data only for some tests, and
5. no disaggregated data (see Table 2).

From 2000–2001 to 2001–2002, the percentage of states that reported disaggregated participation and performance in-

formation for all their general assessments rose from 56% to 70%, that is, from 28 states to 35 states. The number of states that reported performance data but not participation data either for some or all of their tests dropped from 12% to 8%. In general, states that reported disaggregated data for students with disabilities at the state level tended to report disaggregated results at the district and school levels as well.

States that provided information for all of their assessments tended to be ones with both large and small student populations. States that reported disaggregated participation and performance data for their general assessments did so regardless of whether they had one assessment or multiple assessments (18 of the 28 states had more than one assessment in 2000–2001, and 20 of the 35 had more than one assessment in 2001–2002) and regardless of whether they tested in just a few grades or in as many as 10 grades.

In 2000–2001, a total of 10 states reported participation and performance on some, but not all, of their general assessments, and in 2001–2002, this number dropped to 9. Of these states that were missing participation and performance data for some of their assessments, 6 were missing information on only one of their tests in 2000–2001, compared to 7 states in 2001–2002. Of these states, each state administered between two and four assessments.

The numbers of states that publicly reported participation and performance data for their alternate assessments in 2000–2001 and 2001–2002 also are shown in Table 2. In 2000–2001, 13 states publicly reported both participation and performance results for their alternate assessments. Another 10 states reported participation only, and 2 states reported performance only. Thus, only approximately 50% of states reported some kind of information about their alternate assessments. In responding to our mailing, many states indicated that they expected to report results for 2001–2002 after the alternate assessment became more reliable and all the “kinks” were ironed out.

In 2001–2002, 22 states publicly reported both participation and performance results for their alternate assessments. Another 5 states reported only performance data, and 5 more states reported only participation data. Thus, 64% of states reported some kind of information about their alternate assessments. Although this was an increase over the previous year, 36% of states did not publicly report any alternate assessment information.

### General Assessment Participation Results

The ways in which states with publicly reported participation data reported those data in 2000–2001 and 2001–2002 are shown in Table 3. Data for the 2 years are similar but slightly lower in 2000–2001. Of all assessments administered, most provided only a count of the number of students with disabilities or special education students tested. For only a handful of assessments were both the number of students with disabilities tested and the percentage tested reported. In 2000–2001,

**TABLE 3.** Participation Reporting Approaches for Students With Disabilities for General Assessments in 2000–2001 and 2001–2002

Reporting approach	2000–2001	2001–2002
Students tested (%)	12	22
Students tested or not tested (%)	14	29
Students tested (without count; %)	1	8
Students absent (count or %)	4	6
Students not tested (count or %)	7	15
Students exempted or excluded (count or %)	9	8
Students tested (count and %)	11	14
Count (without %)	63	62

*Note.* These data are based on 76 tests in 2000–2001 and 86 tests in 2001–2002.

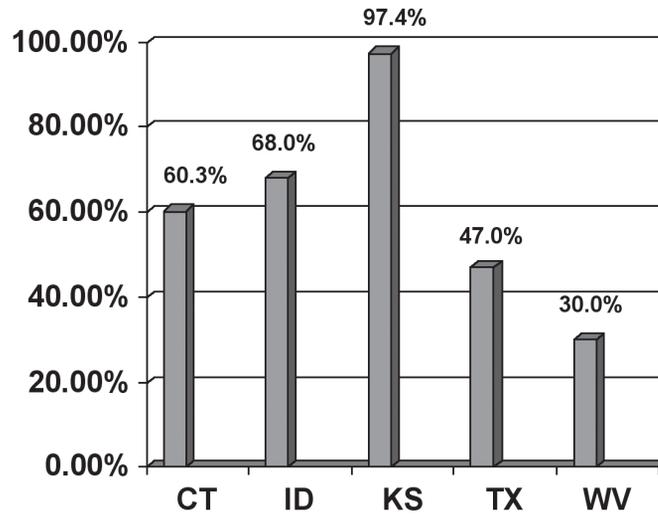
18% of assessments provided a rate (either the percentage tested or not tested); in 2001–2002, that number rose to 34%. Information was sometimes provided about students who were exempted or excluded and students who were absent.

Although the percentage of students with disabilities participating or not participating was reported for 29 assessments in 2001–2002, for example, those assessments came from just 20 states. It appeared that it might have been possible to calculate a percentage tested in several other states because they provided a tested count and an enrollment. However, states used different terms and did not always provide a clear definition of who was represented in their participation results, so we did not calculate results for this study. For instance, the state of Washington reported the percentage of students in special education who were “exempt,” as well as the percentage of students in special education “not tested,” but did not indicate whether those two terms were mutually exclusive. Assuming that they were mutually exclusive would have produced a different rate from assuming that they were not mutually exclusive.

Most states provided participation data according to grade and content area tested. A few states provided more detailed disaggregated data. For 2001–2002, 14 states provided information about students who took an assessment with an accommodation. Of these 14 states, 9 presented only data for students who took the test using accommodations in general. Five states (Arkansas, Colorado, Louisiana, New Mexico, and North Carolina) presented participation and performance data by the specific type of accommodation used (e.g., braille, large print, extended time).

Figure 1 illustrates participation rates reported in those states with clear reporting of participation rate information. We defined *participation rates* as the number of participating students with disabilities divided by the total number of stu-

2000–2001



2001–2002

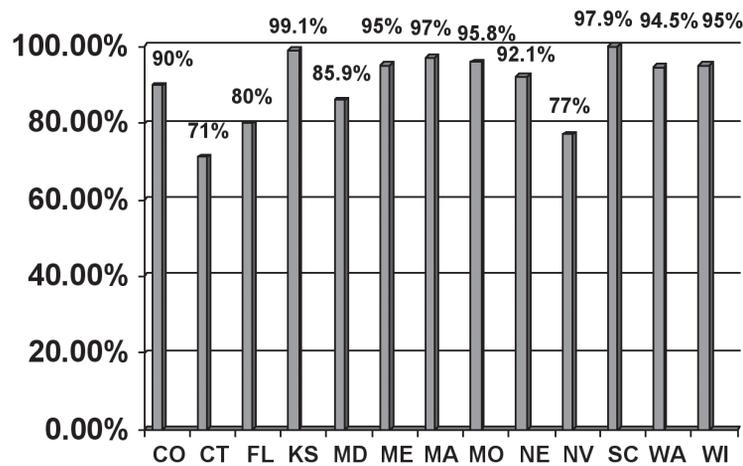


FIGURE 1. Percentages of students with disabilities participating in general assessments in states with clear reporting of participation rates in middle school mathematics in 2000–2001 and 2001–2002.

dents with disabilities. For the purposes of summarizing participation rate information, we selected one grade to portray in Figure 1. In most states, participation in the middle school/junior high math test was used. If the state tested in more than one middle school grade, we used eighth-grade data. We selected this grade because as grade level increases, participation rates tend to decrease; therefore, elementary school rates would be high. We thus decided that eighth-grade rates would be most representative of an average rate.

For 2000–2001 (see Figure 1), only five states had participation rates clear enough to present. In West Virginia, partici-

pation rates for students with disabilities who were considered to have participated in the standard administration of the test included all of those students across 3rd through 11th grades (i.e., data were not disaggregated by grade or subject). Wisconsin also provided a participation rate (68.5% of students with disabilities) for its third-grade reading comprehension test, but because the results were in a different subject and grade level, they are not included in Figure 1.

In 2001–2002, we tightened our criteria for including participation data and did not include data that were aggregated across grades and subjects; this eliminated data from Ok-

lahoma, Texas, and West Virginia. Even with this more stringent rule, 13 states provided a participation rate (see Figure 1). All of the rates were above 70%, and 6 of the 13 states had participation rates of 95% or higher. We used eighth-grade mathematics, as we had done during the previous year, but allowed greater inclusion of content and grades with our targets. For example, in Nevada, the eighth-grade data are combined for the reading and math tests, and in South Carolina, data for all grades are combined for the math assessment. Three states (Illinois, Kentucky, and New Hampshire) provided a participation rate that was the number of students with disabilities who participated out of *all* students rather than a percentage of students with disabilities who were tested. Although this is more helpful than not providing any rate, we did not include this information in Figure 1 because it was a different type of rate from the majority of rates we had available to us.

Participation results in Figure 1 were obtained from the different types of tests that were used in the states. Some states used criterion-referenced tests, whereas others used norm-referenced tests. Several states administered a test that had both criterion and norm-referenced components. In addition, each state had a different way of calculating participation rates. Some states used the percentage of students enrolled on the day of testing; other states used Child Count enrollment data gathered in December.

### *General Assessment Performance Results*

In 2001–2002, 48 states reported performance results for students with disabilities. Thirty-nine states provided these results for all of their assessments, and an additional 9 states gave these results for some of their assessments. This number was up slightly from 44 states in 2000–2001.

When looking at the performance of students with disabilities on states' 2001–2002 criterion-referenced assessments, we found a large range in the percentage of students who were proficient on their state assessments. On several state assessments, extremely low numbers of students were proficient. For example, in one state, 0% of students with disabilities were proficient in middle school math. At the elementary level in reading, 32% of states that reported performance data indicated that less than 20% of their students with disabilities were proficient on their state test. By the high school level, 93% of states had fewer than 20% of students with disabilities scoring as proficient. This pattern was similar for math results. As the grade level of students with disabilities increased, fewer scored as high on their state criterion-referenced test relative to their peers without disabilities. In other words, the gap between students with and without disabilities increased.

When looking at performance on exit exams in 2000–2001, we found that higher rates of students were proficient on these tests than on other high school–level criterion-referenced tests. In reading, of the 12 states that provided performance data for their exit exams, fewer than 20% of students were proficient in only one state. In 5 states (42%), more than 50% of

students with disabilities were proficient. On mathematics exit exams, 3 states (25%) had fewer than 20% of students with disabilities who scored as proficient, and 3 states (25%) had more than 50% of students with disabilities score as proficient.

## Discussion

The data in this study provide a baseline perspective on the participation and performance of students with disabilities in state assessments. The year 2001–2002 was the first year of the 12-year NCLB timeframe for states to have all students reach proficiency. Public reporting is one view of what this baseline is like. Indeed, states were not pushed to report their data immediately to schools and districts so that they would know the results for accountability purposes until 2002–2003. From this perspective, the data presented here are a baseline, but a baseline without the same urgency to get the results out to the public as is now the case for NCLB reporting.

The collection and verification of data for a summary report such as is reflected in this article requires considerable time on the part of the researchers. This time is needed for searching through complicated Web sites; obtaining verification responses from state directors who are already swamped in their day-to-day work of preparing and administering assessments and then calculating their AYP results, along with numerous other tasks; and finally, digging through the data to determine exactly what are the correct numbers to enter and where. Thus, although it will be important to look at the 2002–2003 data against this baseline, it is going to take some time still to have those data all gathered, verified by states, and summarized.

Although 2001–2002 is a baseline for NCLB reporting, both 2000–2001 and 2001–2002 provide a picture of the progress that states are making in relation to the IDEA 1997 requirement that states report disaggregated data on the participation and performance of students with disabilities. Our study indicated that states' public reporting on students with disabilities has improved dramatically in the past few years, especially in comparison to the years prior to this study. In 1999, only 14 states reported on the participation of students with disabilities in assessments, and only 15 reported on the performance of these students (Thurlow, Nelson, Teelucksingh, & Ysseldyke, 2000). Still, despite federal laws that require disaggregated reporting, only 70% of all states reported participation and performance results in 2001–2002 for students with disabilities on *all* of their general state assessments. Alternate assessment information was reported with much less frequency than was general assessment data. Only 32 states reported any alternate assessment data for assessments administered in 2001–2002, a finding that possibly indicates that the development of these assessments took longer than expected in some states or perhaps that it is more challenging for states to get alternate assessment results into the queue for public reporting. Only 22 states reported both participation and performance data for their alternate assessments, compared

to the 35 states that reported both participation and performance data for all of their general assessments. All of the states that reported alternate assessment data also reported general assessment data. Although the amount of information that was reported for the alternate assessment increased from 2000–2001 to 2001–2002, many states still did not report this important information.

When reporting participation data, states reported a count for the majority of their assessments ( $n = 62$ ). For only 22 assessments (26%) was the percentage of students with disabilities tested given, and an additional 7 assessments provided the percentage of students not tested. This means that a total of 29 assessments (34%) provided a rate. This is disappointing, given that rates of participation are the only way it is possible to tell whether the majority of students with disabilities were included in the assessment.

The variability that was evident in the mathematics participation data presented in Figure 1 is reflective of what we saw when we examined the available participation rate data. What these data tell us is that states are likely to have a difficult time meeting the required 95% participation rate required for No Child Left Behind unless something dramatic happens in the period of a year. Of the 13 states with clear participation rates in the reports we examined, only 6 had rates above 95%. Of course, these figures do not add in the percentage of students in the alternate assessments. With these students added in, assuming a number somewhat close to the 1% cap allowed by NCLB for proficiency scores (which would translate to about 10% of students with disabilities), many states are in a much better position. There may still be questions about how the percentages are derived. In other words, exactly what numbers make up the denominator that is used to calculate a percentage? There has been considerable evidence in the past that the numbers that go into percentages may have many questionable assumptions underlying them (Erickson, Ysseldyke, & Thurlow, 1997). States have other adjustments they can make to their percentages (such as averages across 2–3 years), but in some states it is likely that participation rates are still going to be an issue for this subgroup.

When examining performance results, we found a wide range in the percentage of students who were proficient on state assessments. Although a large percentage of the states had between 15% and 30% of their students with disabilities scoring as proficient, several had some very high percentages in the 70s, 80s, and even 90s. This is probably due to differing standards or degrees of rigor among the state tests. The pattern of lower percentages of students with disabilities proficient at higher grade levels nevertheless is consistent across states. By high school, few states had more than 20% of students with disabilities who were proficient on their state criterion-referenced tests.

Ysseldyke and Bielinski (2002) explored this type of growing performance discrepancy between general education and special education students and found that it could be attributed to several possible factors. Many of the factors reflected

changes in the composition of the special education population that are tied to student achievement. For example, every year, many of the higher performing students in special education are returned to general education, losing their special education status, whereas many of the lowest-performing general education students are placed in special education. Contributing factors included the following: (a) special education students with mild disabilities drop out in greater numbers than students with more severe disabilities, (b) high school assessments may be less valid for students with disabilities and be less aligned with the students' curriculum, and (c) high school students are less likely to receive test accommodations than are students in younger grades. All of these factors come together to produce a situation in which the general education population may exhibit performance gains but the special education population falls farther behind.

The challenge of students in the disability subgroup who change over time is not a sufficient explanation of either the gap between students with disabilities and general education students or what might be expected to happen to the gap over time. The Center on Education Policy's (2004) report on the second year of implementation of NCLB documented state education administrators' beliefs about what will happen to the achievement gap over time. For students with disabilities versus students without disabilities, nearly half of the responding state administrators ( $n = 45$ ) indicated that they thought that the "gap will narrow somewhat." This was the response given in the majority of cases for all subgroups.

When looking at performance results on high school exit exams, we found that student performance was higher on exit exams than on high school criterion-referenced exams. This could be because exit exams often are composed of material at the 8th- or 9th-grade levels, compared to criterion-referenced tests, which have more 10th- or 11th-grade material (Achieve, 2004a, 2004b). States need to ensure that students graduating from high school have adequate skills; therefore, high school exit exams should not be too easy. Instead, states should set high standards, and teachers should then work with students with disabilities to help them make continual progress toward those goals.

More time is needed to determine how well states are making progress toward the NCLB goal of having 100% of students with disabilities scoring as proficient by 2013–2014. Although some states are clearly on their way toward high rates of student proficiency, almost half the states had less than 10% of their students with disabilities scoring as proficient for at least one grade level. This wide performance distribution illustrates the challenging work ahead for many states. Very little research has been conducted to help explain the reasons for this gap. Still, the survey of states reported by the Center for Education Policy (2004) acquired some responses that suggested that reporting was seen as one avenue for beginning to address the gap issue. For example, the Center's report included this statement from a respondent: "The disaggregation of data by various subgroups will bring attention to

achievement gaps” (p. 16). Similarly, another response was that “the data collection and reporting requirements are challenging to meet but will provide a needed focus on the learning of all students that has been masked in the past” (p. 16). (This quotation went on to identify several problems with NCLB.) Reporting, while contentious, is clearly a mechanism for making gaps evident, including the gap between special education students and general education students.

The results of our study suggest that states’ disaggregated reporting is still not all that it should be. We offer three recommendations for improving reporting practices. First, states should provide data in a timely manner, ideally no more than 6 months after the administration of a test. Our analyses revealed that some states reported their data much more quickly than did other states. Of course, a wide range of variables affects the speed with which data can be scored, analyzed, and reported back to states, districts, and schools, including the number of students tested and the state’s financial resources to support a quick turn-around. The recent requirements of NCLB, which call for timely reporting of assessment results, makes the need for timely data summaries and disaggregations even more critical.

Second, states should establish reporting practices that are consistent with IDEA ’97; therefore, data should be disaggregated for all assessments for which the performance of general education students is reported. Participation and performance results should also be reported for students taking their state’s alternate assessment. Although IDEA requires that states report participation and performance for students with disabilities in the same detail and frequency as for students without disabilities, our analysis found that only 35 states reported both participation and performance rates for students with disabilities on all the assessments for which information was provided about students without disabilities. Of course, these data were for the 2001–2002 school year, which is 1 year prior to the annual yearly progress requirements of NCLB. Still, the reporting requirements actually were in place for IDEA, regardless of whether states perceived those requirements to have the same force as the NCLB requirements.

Third, states should report participation *rates* for students with disabilities (U.S. Department of Education, 2003), ideally, based on test-day enrollment (Ysseldyke et al., 1998). When states provide the number of students with disabilities who were tested, the information is less helpful than when the number is presented along with the percentage of students tested. Providing a rate makes it easier to obtain a picture of how many students are participating in general assessments and in alternate assessments. It is also important to explain who is being included in the count. When presenting the percentage of students with disabilities who were tested, some states used the entire student body as the denominator, whereas others calculated the rate using the total number of students with disabilities as the denominator. States must specify which students they are using as the denominator and use a count based on the enrollment on the day of testing, if possible (Erickson,

Ysseldyke, & Thurlow, 1997). If that count is not available, the next closest count should be used. This may be the December 1 Child Count for students with disabilities.

The guidance that has been available to states has been limited, and states have been left to their own devices to figure out how to report on the participation and performance of students in their assessments. When states did this initially, they usually were not worried about subgroups of students. For the most part, states began retrofitting their reporting systems to include information on students with disabilities when required by IDEA. With the advent of the specific and detailed reporting needs of NCLB (U.S. Department of Education, 2003), retrofitting may not be enough. States now need to report the percentage of students tested (or the percentage not tested), the proficiency level of all students, and the most recent 2-year trend data on student achievement for each grade and subject. This must be done for all students and subgroups, including students with disabilities.

Although NCLB will lead states to begin reporting more detailed assessment data, we need to continue to watch the evolution of reporting and the effects of public accountability on students with disabilities. It will be essential to push for transparency in reporting so that the public can determine how states are handling some of the challenges in assessing students with disabilities. For example, are scores from nonstandard test administrations reported and, if so, how and where? Do students who use certain accommodations count as participants but then disappear from performance calculations (and accountability measures)? Do the subgroup size allowances built into NCLB result in underreporting of students with disabilities at the school and district levels, the levels at which educators need to examine the data most closely? To what extent do the reporting requirements result in blaming subgroups like students with disabilities for not making adequate progress rather than the targeting of resources with expectations of success? The possibilities of public reporting are great for the disability subgroup, but so are the perils.

#### AUTHORS’ NOTE

Preparation of this article was supported in part by a cooperative agreement (H326G000001) with the U. S. Department of Education, Office of Special Education Programs. Points of view or opinions expressed do not necessarily reflect those of the USDOE or offices within the department.

#### REFERENCES

- Achieve. (2004a). *Do graduation tests measure up? A closer look at state high school exit exams*. Washington, DC: Author.
- Achieve. (2004b). *Ready or not: Creating a high school diploma that counts*. Washington, DC: Author.
- Bielinski, J., Thurlow, M. L., Callender, S., & Bolt, S. (2001). *On the road to accountability: Reporting outcomes for students with disabilities* (Technical Report 32). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

- Center on Education Policy. (2004). *From the capital to the classroom: Year 2 of the No Child Left Behind Act*. Washington, DC: Author.
- Cibulka, J. G., & Derlin, R. L. (1995). *State educational performance reporting policies in the U.S.: Accountability's many faces*. (ERIC Abstract ED 401 613)
- Erickson, R. N., Thurlow, M. L., & Thor, K. (1995). *1994 state special education outcomes*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Erickson, R., Ysseldyke, J., & Thurlow, M. (1997). Neglected numerators, drifting denominators, and fractured fractions: Determining participation rates for students with disabilities. *Diagnostique*, 23(2), 105–115.
- Fast, E. F., Blank, R. K., Potts, A., & Williams, A. (2002). *A guide to effective accountability reporting*. Washington, DC: Council of Chief State School Officers.
- Koenig, J. A. (2002). *Reporting test results for students with disabilities and English-language learners: Summary of a workshop*. Washington, DC: National Academy Press.
- Koretz, D., & Hamilton, L. (2001). *The performance of students with disabilities of New York's revised Regents Comprehensive Examination in English* (CSE Technical Report 540). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- McGrew, K. S., Thurlow, M. L., Shriner, J. G., & Spiegel, A. N. (1992). *Inclusion of students with disabilities in national and state data collection programs* (Technical Report 2). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- McGrew, K. S., Thurlow, M. L., & Spiegel, A. N. (1993). An investigation of the exclusion of students with disabilities in national data collection programs. *Educational Evaluation and Policy Analysis*, 15, 339–352.
- National Research Council. (1999). *Testing, teaching, and learning: A guide for states and school districts*. Washington, DC: National Academy Press.
- Shriner, J. G., & Thurlow, M. L. (1992). *State special education outcomes 1991*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Thompson, S. J., Thurlow, M. L., & Lazarus, S. (2001). *Reporting on the state assessment performance of students with disabilities* (Topical Review No. 3). College Park: University of Maryland, Education Policy Reform Research Institute.
- Thurlow, M., Elliott, J., & Ysseldyke, J. (1998). *Testing students with disabilities: Practical strategies for complying with district and state requirements*. Thousand Oaks, CA: Corwin Press.
- Thurlow, M. L., House, A., Boys, C., Scott, D., & Ysseldyke, J. (2000). *State participation and accommodations policies for students with disabilities: 1999 update* (Synthesis Report 33). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M. L., Langenfeld, K. L., Nelson, J. R., Shin, H., & Coleman, J. E. (1998). *State accountability reports: What are states saying about students with disabilities?* (Technical Report 20). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M. L., Nelson, J. R., Teelucksingh, E., & Ysseldyke, J. E. (2000). *Where's Waldo? A third search for students with disabilities in state accountability reports* (Technical Report 25). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M., Quenemoen, R., Thompson, S., & Lehr, C. (2001). *Principles and characteristics of inclusive assessment and accountability systems* (Synthesis Report 40). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M. L., & Thompson, S. J. (1999). District and state standards and assessments: Building an inclusive accountability system. *Journal of Special Education Leadership*, 12, 3–10.
- Thurlow, M. L., & Wiley, H. I. (2004). *Almost there in public reporting of assessment results for students with disabilities* (Technical Report 39). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M., Wiley, H. I., & Bielinski, J. (2003). *Going public: What 2000–2001 reports tell us about the performance of students with disabilities* (Technical Report 35). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M. L., Ysseldyke, J. E., & Silverstein, B. (1995). Testing accommodations for students with disabilities. *Journal of Remedial and Special Education*, 16, 260–270.
- U.S. Department of Education. (1999). *Peer reviewer guidance for evaluating evidence of final assessments under Title I of the Elementary and Secondary Education Act*. Washington, DC: Author.
- U.S. Department of Education. (2003). *Report cards: Title I, Part A: Non-regulatory guidance*. Washington, DC: Author.
- Ysseldyke, J., & Bielinski, J. (2002). Effect of different methods of reporting and reclassification on trends in test scores for students with disabilities. *Exceptional Children*, 68, 189–200.
- Ysseldyke, J. E., Thurlow, M. L., Langenfeld, K., Nelson, J. R., Teelucksingh, E., & Seyfarth, A. (1998). *Educational results for students with disabilities: What do the data tell us?* (Technical Report 23). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Zlatos, B. (1994). Don't test, don't tell: Is "academic red-shirting" skewing the way we rank our schools? *The American School Board Journal*, 181, 24–28.