



FAQs: Measuring Status and Change in NAEP Inclusion Rates of Students with Disabilities

About This Study

What is the purpose of this study?

The purpose of this study is to measure the change in state-level NAEP inclusion rates while taking into account the differing demographics and inclusion policies in each state. The goal is to report on change in inclusiveness for all states and the District of Columbia. While the focus of the report is on measuring change over 2007–09, changes over 2005–07 and 2007–09 were also calculated.

Why was the study conducted?

This study is part of ongoing research by NCES into the inclusion of students with disabilities (SD) in NAEP and is an update of a previous study, [Measuring the Status and Change of NAEP State Inclusion Rates for Students with Disabilities](#) (NCES2009453), using newly released 2009 NAEP data and an updated methodology.

Reporting of trends requires consistency in inclusion practices across years, and the lack of consistency in the inclusion of students with disabilities has been a concern for NAEP researchers. NAEP has repeatedly shown that inclusion rates of students identified as having disabilities vary among the states. In the 2009 grade 4 NAEP mathematics assessment, the national average inclusion rate of students with disabilities who were not English language learners was 85 percent, and state inclusion rates ranged from 70 percent to 94 percent. In the 2009 grade 8 NAEP mathematics assessment, the national average was 78 percent and state inclusion rates ranged from 45 percent to 92 percent. Numerous publications and working papers related to the inclusion of students in NAEP have been conducted and are available on the [NCES website](#).

In July 2005, the U.S. Government Accountability Office (GAO) released a report titled *No Child Left Behind Act: Most Students With Disabilities Participated in Statewide Assessments, But Inclusion Options Could Be Improved*. In the report, the GAO recommended that NAEP “work with the states, particularly those with high exclusion rates, to explore strategies to reduce the number of SD students who are excluded from the NAEP assessment.”

NCES responded with the following actions:

- Researched the local decision-making process for participation by and accommodation of students with

disabilities in NAEP;

- Implemented a process to determine whether students could participate in NAEP without their usual accommodations if not permitted on NAEP; and
- Improved training of NAEP administrators and field staff for 2007 and subsequent assessments, which clarified the process for increasing inclusion in NAEP.
- NCES also conducted research (Kitmitto and Bandeira de Mello 2009) to develop a methodology for measuring state inclusion rates while taking into account the differing demographics and inclusion policies in each state. The current study is a continuation of that previous research and development study.

The Data and Population Analyzed

What assessments were analyzed?

This report analyzed 2005, 2007, and 2009 assessments in mathematics and reading, in grades 4 and 8. It focuses on state-level changes in inclusion rates from 2007 to 2009. Changes from 2005 to 2007 were also analyzed given that the statistical model used differed slightly from that used in the previous report for analyzing inclusion rates for 2005 and 2007.

How is a student with a disability defined?

NAEP's definition of a student with a disability includes students with an Individualized Education Plan (IEP) for reasons other than being gifted or talented, students with a Section 504 Plan, and any other students who have received an accommodation on NAEP.

What population is analyzed in this study? Why are English language learners not included?

This report focuses on the inclusion of SD students who are not English language learner (ELL) students. ELL students were not part of this analysis because the factors influencing the inclusion of SD and ELL students are distinct.

The inclusion of students with disabilities on NAEP

Why would a student not be included on NAEP?

A disabled student may be excluded from the NAEP assessment if the student has a significant cognitive disability that prevents the student from being able to meaningfully participate or access the NAEP assessment with any of the allowed accommodations. The Individuals with Disabilities Education Act (IDEA) requires that

all SD students participate in state-wide assessment programs, with appropriate accommodations when necessary, or through alternate assessments. The Elementary and Secondary Education Act (ESEA) also requires the participation of SD students, as well as ELL students, in the academic assessments required under that Act. Although federal law does not explicitly specify similar requirements regarding the participation of SD and ELL students in NAEP, the NAEP program has been working to make its sample of students who take the assessments as representative as possible of all students.

Do state assessments and NAEP provide the same accommodations for students with disabilities?

State assessments may allow [accommodations](#) for SD or ELL students that would not be allowed on NAEP. Because the requirements for state assessments are not the same as for NAEP, one would not expect the inclusion rates to be the same. For reading, examples of accommodations allowed by states but not NAEP include having test items read partially or wholly to the student, and using a dictionary, thesaurus, or spelling/grammar-checking software or devices. For mathematics, one example is using a calculator. An example of a general accommodation not allowed on NAEP but that might be used on a state assessment is taking the test over multiple days. Unlike many states, NAEP does not offer an alternate assessment.

What determines whether or not a student with a disability is included in NAEP?

For every student selected to participate in NAEP and identified as an SD student, a questionnaire is intended to be filled out by the special education teacher or staff member who is most familiar with the student. This questionnaire leads the school staff member through a decision process to determine whether the student should be included in the NAEP assessment without an accommodation, included with an allowed accommodation, or not included. See copies of the questionnaire.

What is a state's "inclusion rate of students with disabilities" or "actual inclusion rate?"

A state's inclusion rate of SD students is the weighted percentage of SD students in the state sampled by NAEP who participate in NAEP. In other words, the weighted number of SD students in a state who are selected for participation in NAEP is in the denominator, the weighted number of those students who participate in NAEP is in the numerator, and the fraction is multiplied by 100 to turn it into a percentage. NAEP has repeatedly shown that inclusion rates of students identified as having disabilities vary among the states.

This inclusion rate is referred to in this report as the state's "actual inclusion rate."

What is the target inclusion rate for a state? What should a state's inclusion rate be?

This report makes no claim to have determined what any state's inclusion rate should be. Averages are used to set benchmarks for prediction and measurement, but these are not to be interpreted normatively. In the spirit of IDEA and NCLB, however, higher inclusion rates are considered better than lower inclusion rates.

Factors associated with a state's inclusion rate

Why might a state's inclusion rate of students with disabilities change?

Students with less severe disabilities, such as a speech or hearing impairment, are more often included in NAEP testing. Students with more severe disabilities, such as mental retardation, are less often included in NAEP. One expects a state's inclusion rate to change due to changes in the distribution of the characteristics of its SD students. The characteristics that can be identified for use in this study are type of disability (learning disability, speech impairment, mental retardation, emotional disturbance, autism, other health impairment, and other disabilities), severity of disability (severe, mild, moderate), whether the student has multiple disabilities, whether the student has an Individualized Education Plan (IEP), and whether the student receives an accommodation on his or her state assessment that is not allowed on NAEP. A state's inclusion rate may also change due to changes in NCES policy and practices, state efforts to include more students, and other factors.

Do differences in accommodation policies across states enter into the analysis?

This is an important factor that is taken into consideration by an indicator for when a student receives an accommodation on the state assessment that is not allowed on NAEP. Other differences between state policies are not controlled for and any impact they may have on inclusion rates would thus be captured in the change measure.

What about differences across states in accommodation rates on NAEP?

The concern of this study is only focused on participation in NAEP. Accommodations allowed on NAEP are considered permissible ways to increase the participation of SD and ELL students. Hence, there is no attempt to control for changes in accommodation rates on NAEP because the accommodations are legitimate ways for states to become more inclusive. Controlling for a factor means that change due to that factor will not be captured in the "change" measure. Since changes in accommodation rates on NAEP are not controlled in this methodology, if such changes lead to changes in inclusiveness in a state, that change in inclusiveness will be captured in the "change" measure.

Measuring the change in inclusion rates and the status of inclusion rates

How does the report measure a state's "change" in inclusiveness?

Part of a state's actual change in inclusion rates is explained by shifts in the characteristics of the state's SD population and part is due to other factors. The portion that is not explained by shifts in a state's SD population characteristics is considered a measure of "change in inclusiveness."

The methodology used for partitioning the actual change into a portion that is explained and a portion not

explained is derived from a technique called the Oaxaca-Blinder decomposition and is detailed in the previous report, [Kitmitto and Bandeira de Mello \(2009\)](#).

The methodology is as follows: First, student-level benchmarks of inclusion (probability of inclusion) were estimated for each profile of student characteristics based on relationships found using 2005 data. Second, a state-level benchmark of inclusion (predicted rate of inclusion) for a state in any given year was estimated by averaging the student-level benchmarks for the various types of students with disabilities in that state. Finally, change in inclusiveness was measured across time in relation to these benchmarks.

This report uses two different approaches for measuring change in inclusion rates: a nation-based and a jurisdiction-specific approach. The nation-based approach uses national averages to set benchmark inclusion rates for each type of student. The jurisdiction-specific model, an alternate approach, uses averages in each state to set benchmark inclusion rates for each type of student for that state. The primary measure used in this report is the nation-based measure while the jurisdiction-specific approach is used to check the robustness (stability) of the nation-based results.

How do the nation-based and jurisdiction-based approaches differ?

The nation-based and jurisdiction-specific approaches differ in how the predicted probabilities of inclusion (student-level benchmarks) are determined for each student based on his/her demographic characteristics. The nation-based method uses estimates of the 2005 national average inclusion rate for students with the same background characteristics as the predicted probability. The jurisdiction-specific method, on the other hand, uses estimates of the 2005 state average inclusion rate for students with the same background characteristics as the predicted probability.

Which approach is better, the nation-based or the jurisdiction-based approach?

Each approach has its advantages, and the results have been found to be highly correlated. The nation-based method has the advantage that characteristic categories can be used jointly to identify more distinct “types” of SD students. This is because the student-level predicted probabilities of inclusion (student-level benchmarks) are derived from a 2005 national dataset that combines all of the state datasets. The jurisdiction-specific approach derives student-level predicted probabilities using only a given state’s 2005 data. The jurisdiction-specific model has the advantage that the predicted probabilities used are particular to that state and will reflect practices and policies in that state. It is important to remember that these are simply two different ways of setting benchmarks and the resulting change measures essentially capture the same thing: how a state has changed in inclusiveness.

What about changes over time in the identification rates of students with disabilities?

In this report, changes in identification rates of SD students are not explicitly accounted for, but changes in the average type of student considered SD are accounted for through the use of control variables such as disability type and level of severity. For example, if more students with less severe disabilities are considered SD in a state, these students will have high expectations for inclusion and this will raise the state’s benchmark. Year-to-year changes in the identification policy used in a state can cause some inaccuracies in the method, but the use of control variables is expected to minimize these.

How does the “status” of a state’s inclusion rate in an earlier year affect its status in a later year?

States that include SD students at high rates initially have less potential for increasing their inclusion rates. Hence, there is less expectation that such states will increase inclusion rates. For interpretation of a state’s change in inclusiveness, it is important to consider a state’s relative inclusiveness in the initial period as a context for understanding change. The status of a state’s inclusiveness is made in comparison to other states’ rates in the starting year of the period over which change is being measured.

How is the “status” of inclusiveness calculated?

The status of a state’s inclusiveness is measured in the first year of the span of time over which change is measured. Similar to measuring change in inclusion rates where differences in a state’s SD population across time are controlled, the differences in SD populations across states also need to be controlled when measuring status. The status measure is the difference between a state’s actual inclusion rate and its benchmark inclusion rate. The benchmark inclusion rate is the inclusion rate predicted by the characteristics of the state’s SD population. States whose actual inclusion rate is higher than its benchmark are relatively more inclusive than other states.

How is the status measure used?

The status measure is a number on a continuous scale that compares inclusiveness in a state to the average inclusiveness in the nation were all states to have an SD population with similar characteristics. This report uses a method of presentation that simplifies the results to help readers easily grasp the information. In the simplified presentation, the status measure is used to group states into four categories by level of inclusiveness, with roughly equal numbers of states in each group. These are quartiles of inclusiveness. The first quartile contains states with the lowest status measures and the fourth quartile contains states with the highest status measure. The change measure is also simplified by placing states into three categories according to the direction and significance of their change measure (decrease, no change, increase). The three categories of change are not required to have equal numbers, and a category could potentially contain no states at all. The simplified groupings (four for status and three for change) are used to place states into 1 of 12 cells in a 3x4 grid.

Benchmarks

What are “student-level” benchmarks?

A student-level benchmark is a predicted probability that a student with given characteristics will be included in NAEP. Students with characteristics associated with higher inclusion rates (such as those with a specific learning disability or those with a mild disability) have a higher benchmark for inclusion, and students with characteristics associated with lower inclusion rates (such as those with mental retardation or those with a

severe disability) have a lower benchmark. The relationships between student characteristics and benchmarks are calculated from the estimated parameters of the statistical model used in this study. The parameters are estimated using 2005 NAEP data.

Student-level benchmarks differ by student characteristics but they do not differ across time. In other words, for a given profile of student characteristics, the student-level benchmark will be the same for such students in 2005 and 2007 and 2009. Suppose, for example, the model estimated that students with a specific learning disability that was mild and who had an Individualized Education Plan (IEP) and did not receive an accommodation on the state assessment that was not allowed on NAEP were included 90 percent of the time using 2005 data. This would be the benchmark for that type of student. In all years and in all states, students of this type would be expected to be included 90 percent of the time.

What are “state-level” benchmarks?

A state-level benchmark is an aggregation of its students’ individual-level benchmarks. By averaging student-level benchmarks to the state level, a state’s benchmark takes into consideration the characteristics of its students. In this manner, the differing populations of students with disabilities across states and across time lead to different state-level benchmarks for measurement. While the benchmark for any given student profile does not change across time, if the distribution of student profiles in a state changes, the benchmark for that state will differ across time.

Miscellaneous

How are standard errors calculated?

The report provides standard errors and significance testing for the change measure, as that measure is the focus of this report, but not for the starting point measure. To calculate the standard error, a modification of NAEP’s recommended process is used. The modification is necessary because the analysis uses two NAEP administrations to calculate results, and error from both data sources needs to be combined. See the [2009 R&D report](#) for more details on how standard errors were calculated.

What explains why a state has a high or low change measure?

Answering this question is beyond the scope of this report. NCES policy and practices and state efforts to include more students are the major forces for change in inclusion rates that are expected to be captured in the change measure, but it is not known which of these might be driving results, and it is possible that other factors might be at work as well.

Are there any caveats in interpreting these findings?

First, it is very important to be clear that the student-level benchmarks (predicted probabilities of inclusion) and state-level benchmarks (predicted inclusion rates) presented in the report are not to be interpreted prescriptively. This report makes no claims as to what a state’s inclusion rate should be; it uses averages to

set benchmarks for prediction and measurement. Second, there is probably some error in the measurement of student characteristics (type of disability, severity of disability, accommodations on the state assessment). The consequence is that the greater the measurement error, the smaller the amount of the actual change that will be explained by changes in the SD characteristics, and, hence, the larger the amount of change that will be captured in the change measure. Third, different student characteristic variables may be defined differently in different states. This will cause bias in the nation-based measure of change but will not be a problem in the jurisdiction-specific method.

