Towards a Better National Assessment Program

NSW Teachers Federation
Submission to the NAPLAN Review

Content

Overview

The teaching profession continues to recognise that it is essential to have a National Assessment program for Australia's students. Such a program is imperative if we are to support communities most in need, to track how educational standards are developing and to assist individual students to grow and progress to their optimal level.

The existing National Assessment Program fails in each of these criteria.

This paper will spend only a little time outlining the profound shortcomings of the existing National Assessment Program. Its problems are manifest. The debacle of the 2019 operation was only the most recent example of its futility and its corrosive effects.

The existence of this current review is recognition, in itself, that thorough ongoing change is essential and urgently needed and a welcome beacon that opportunity is at hand.

It would be a tragedy to waste that opportunity.

The teaching profession accepts its responsibility to assess our students, and teachers are constantly using processes of formal and informal assessment of the students in their care. The goal is, as ever, to ensure that assessment focuses on the best interests of young people and builds their capacity as learners.

The National Assessment Program, as it stands, has failed by the standards it has set itself. It has not contributed to an increase in educational outcomes. It has heaped public scorn on disadvantaged students and communities, which are placed in the modern day stocks through the invasive My School website. It rewards a narrow band of often lower-order intellectual capacities; it has narrowed the taught curriculum; it has corresponded to a seemingly inexorable decline in Australia's performance in major international tests.[1]

The National Assessment Program as it stands has also generated bogus comparisons between children being tested in different modes; it has contributed to unwarranted stress in children; it has invaded the privacy of children and communities; it has diminished respect for the higher purposes of schooling; it has reduced children and schools to mere data sets for the gaze of ignorant media commentators; it has taken teachers and students away from their regular teaching and learning programs for significant periods of time, and most recently, it has made Australia a global laughing stock with its most recent contradictions and national disruptions.

We concur with the view of John Collier, head of St Andrew's Cathedral School, when he observes: "I don't know of any educator who would advocate for NAPLAN as what it's become."[2]

Our children and our nation deserve so much better.

A basis for a new national assessment program framework

The existing National Assessment Program is now burdened with a vast range of ascribed purposes and policy tasks.

Over the years, NAPLAN has been given the roles of:

• a diagnostic test of literacy and numeracy across primary and secondary education

• a diagnostic test of how an individual child is progressing

• a measure of an individual teacher's performance

• a measure of a school's overall performance

• a measure of the effectiveness of a particular state or territory's performance

• a measure of a school's effectiveness over time

• a device used to compare schools and systems

• an instrument of comparison between so-called "like" schools

• subject matter for dinner parties

• a device to be deployed by parents when choosing schools

• a detection measure to identify schools performing highly under the NAPLAN regime

• a detection measure to identify schools "coasting" under the NAPLAN regime

• a detection measure to identify schools "underperforming" under the NAPLAN regime

• a source of information for the media on school, system, state and national performance in any given year and over time

• a databank for media and political commentators to use when forming views about contemporary education

• a databank to be mined at the My School website when used for any or all of the above purposes.

The list above is not exhaustive, but is already absurdly long, complex and contradictory.

> *This paper takes the position that the effectiveness of an assessment program is inversely proportional to the number of tasks assigned to it.*

This paper asserts that a new National Assessment Program should have two tasks, both of which focus on the needs of children.

It is proposed that a new National Assessment Program requires the following elements:

1. A system of national assessment, regularly conducted, based on a huge, comprehensive and inclusive sample of students deploying a full range of levels of intellectual challenge that provides parents, teachers and public officials with a clear understanding of how various social groups, jurisdictions, and parts of the country are progressing. Initially these assessments would concentrate on literacy and numeracy but may expand over time to assess other parts of the curriculum.

   This would represent Australia's report card.

2. All schools would be additionally required to regularly assess how their students in identified years are progressing by undertaking on-call assessments, supplied from a national bank of test items that are aligned to the Australian Curriculum. This on-call testing would be on focus areas determined by the respective schools and teachers and based on the specific program of teaching experienced by children in their classrooms. It is expected that literacy and numeracy would be given

prominence in these on-call assessments but other learning areas will also be developed. The results of these tests would be individually available to parents and caregivers of students and give a picture of how well a child was progressing in areas actually taught in their classroom and also provide a progress report of how a child was growing in relation to the national standards.

This component relies on the simple but profound principle of *testing to the teaching* rather than the current unhappy state where so many schools feel compelled to teach to the test.

The essence of the paired testing system is to provide valid and reliable evidence of both how the nation is advancing and how each individual child is progressing.

Australia can learn a great deal in the development of these paired testing processes from the best, cognate, international practice and this paper will outline these global exemplars. This paper will outline significant intellectual precedent that provides the research evidence base for all of the proposals herein, gathered from Singapore, the United States, Scotland and New Zealand.

This paper will elaborate on each of these proposals and provide the argument and research evidence in support of each of the propositions.

Australia's report card

Australian education can achieve all of its national assessment goals through a paired system of country-wide testing and individual on-call tests.

The national testing can be most effectively implemented through a mass, regular, comprehensive and inclusive assessment

system that uses world's best practice in sample testing.

Put simply, we can glean the necessary information through a rigorous, valid and reliable process where a representative proportion of students are tested rather than having the census testing that currently prevails and which interrupts the schooling of children in their millions.

Tim Dodd in the *Australian Financial Review* reached the same conclusion:

> "There's no need to test all of Australia's 3.8 million students on a regular cycle to get this information. We could test a far smaller sample."[3]

A national sample size of 100,000 students with an appropriate sampling frame and weighted sample elements could achieve everything the current NAPLAN purports to deliver in terms of tracking the nation's progress and locating areas of educational need. Effective sampling must deploy the selection of appropriate data characteristics to provide a representative profile of the nation as a whole with a sample size large enough and inclusive of all of the elements one wishes to identify and evaluate. This sample might include a constant representative subset maintained over time in order to provide systems and policymakers with further valid longitudinal data to ascertain relevant trends in student and group performance.

As it is continued over time, then progress or regress for communities and social groups can be located, trends addressed and good practice highlighted.

Happily, there are already national and international precedents that establish the efficacy of this approach. The Disadvantaged Schools Program in NSW used such a method for many decades to locate those school communities most in need and which consequently gained additional resource support. The Disadvantaged Schools Program sample survey identified the 500-plus neediest school communities for extra funding

and assistance. It had thorough credibility across the entire education community. The Disadvantaged Schools Program survey worked on the understanding that educational need is a socially and geographically located phenomenon. What's needed is to identify the communities most in need rather than providing massive disruption of a census test of 3.8 million children.

Indeed, global assessment systems such as Program for International Student Assessment (PISA), Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS), oft cited by educators and public figures as indicators of our national trends, have judiciously used sample testing to draw national and international policy conclusions. And each of these testing systems operates on the basis of sample sizes far smaller than what is proposed in this paper.

National Assessment of Educational Progress

It cannot always be suggested that the United States provides us with a profound exemplar in education but with the case of their National Assessment of Educational Progress (NAEP) there is indeed a notable exception.

Their NAEP has been a successful component of their educational landscape for decades now and it is strongly endorsed across the spectrum of their educational community.

Hilary Persky[4] from the Educational Testing Service describes its role and function:

"NAEP's mandate is to report to the nation, states and some districts, sub group assessment results for students, as well as relationships between performance and responses to demographic and subject survey questions.

NAEP is congressionally mandated and is funded by the US Department of Education. An independent oversight group, the National Assessment Governing Board sets policy

for NAEP, determining what subjects will be assessed and working with many hundreds of subject experts to develop frameworks for assessment content.

Once frameworks are complete, they are handed off to test developers for assessment creation. Assessment tasks (and scoring [marking] criteria), in their various iterations from initial development to administration, are reviewed by both subject and measurement experts to be sure they meet framework expectations."

It is useful to note that Australia's educational infrastructure, including an Australian Curriculum, and a well-developed system of national dialogue and jurisdictional cooperation provides an advanced basis for the implementation of an assessment regime partially inspired by NAEP principles.

NAEP is fastidious and most transparent about its sampling methodology. It provides details about how its national, state and district samples are selected and how they are representative of the nation and its constituent elements.

Consequently, NAEP can report for different demographic groups including gender, socioeconomic status and race/ethnicity. It also uses the sampling opportunity to provide fine grain information about the students undertaking the tests, which provides contextual understanding of the subsequent test results.

Because of its multi-layered approach, various levels of government can use the results to compare the outcomes of their students with other states or districts or see how their students are performing against national standards. Trend data is also available and can now be compared over many decades.

NAEP releases reports that focus on key policy areas including achievement gaps for various social groups including summaries of black-white, Hispanic-white, Alaska native and Native American performances. A vital element of the success of the NAEP program is its deep commitment to transparency. A parent, teacher, policymaker or citizen can access

fundamental and comprehensive knowledge about the testing process, the validity of the sampling method and the intentions and uses of all aspects of the NAEP.

One can go to *nces.ed.gov/nationsreportcard/* and use it as a springboard to a deep knowledge of what NAEP seeks to achieve. One can discover popular and technical papers on the sample design, English and multilingual explanations of the program for parents, results, test details and future intentions.

The parent leaflet contains a range of media commentary arising from NAEP that includes:

> "Today's NAEP results confirm that we have a crisis on our hands when it comes to civics education." (*The New York Times*)

> "NAEP is the gold standard of student assessments." (*The Washington Post*)

> "NAEP is the definitive source for state to state comparisons in educational achievement." (*Education Week*)

> "Very few students have the advanced skills that could lead to careers in science and technology according to the results of [NAEP]." (*Lawrence Journal-World*)

> "Urban school districts have steadily increased their scores since 2003." (*The Miami Herald*)

What is clear is NAEP does not quash discussion of educational trends and practices and indeed it is the basis of national, state and local news stories across the United States.

What the National Assessment of Educational Progress does provide is high quality, valid and reliable assessments in order to have those discussions, but without harming any individual child or community.

An additional advantage is that it provides this essential knowledge without massively disrupting regular schooling. NAEP works on a philosophy of low impact testing that causes minimal interruption to teaching and learning programs.

This is in stark contrast to the huge impact that NAPLAN testing currently has on regular schooling.

Leaving aside the many weeks and months that some schools feel obliged to engage in preparing for NAPLAN testing, the high impact of administration, planning, re-rooming, loss of classroom technology facilities, staff training modules, communications, equipment setup, reminders, creation of codes, teacher packs, laptop preparation, re-rostering, multiple lunchtimes, catch-ups, loss of lesson time for teachers' regular classes, collapse of days of the normal school timetable and random unforeseen debacles all detract from a student's right to a full teaching program and a full school year.

NSW school principals surveyed for the Department of Education's Reducing Administrative Burden project reported that NAPLAN takes out at least one full week of teaching time for each of their students. Every year. Every school. Every student.

One can infer that moving to low impact, sample testing would add on average one week's teaching time for every student in the land.

An additional boon would be that the many millions of dollars that are spent across Australia to deliver the current NAPLAN census test regime could be redistributed to fund the improvements in assessment that this review now has the chance to create.

On-call assessments

A fundamental element of a new National Assessment Program would be the continuation of the principle of having externally referenced assessment of students' progress based on programs of instruction that they have actually undertaken. These assessments would be set within the Australian Curriculum in order to establish a comparable understanding and external reference of student development and growth while also providing meaningful information that can be aggregated by policymakers about areas of strength and need emerging within the population.

This component of the proposed National Assessment Program is inspired by Dylan Wiliam's observations about these more formal assessment processes.

Wiliam believes:

"The challenge, therefore, is to create an assessment that is externally referenced, distributive and cumulative. The assessment system needs to be externally referenced [his emphasis], so that the teachers can honestly say to the student, 'These are not my standards'. When the authority for the standard of achievement that students have to reach does not come from the teacher, the teacher is free to be a coach, rather than a judge and jury."[5]

Naturally, such on-call assessment would only be conducted within a sound set of protocols for reporting and not reproduce some of the corrosive and debilitating effects of the existing NAPLAN/My School regime.

Further, they are to be understood within the much broader recognition of what constitutes truly useful assessment.

The statutory board in NSW, the NSW Education Standards Authority, has an excellent starting point here with their Principles of Effective Assessment which states:

"Teachers should consider the effect that assessment and feedback have on student motivation and self-esteem, and the importance of the active involvement of students in their own learning.

Assessment:

- provides opportunities for teachers to gather evidence about student achievement in relation to syllabus outcomes

- enables students to demonstrate what they know and can do

- clarifies student understanding of concepts and promotes deeper understanding

- provides evidence that current understanding is a suitable basis for future learning.

Assessment activities should:

- be based on syllabus outcomes

- be a valid instrument for what they are designed to assess

- include criteria to clarify for students what aspects of learning are being assessed

- enable students to demonstrate their learning in a range of task types

- be reliable, measure what the task intends to assess, and provide accurate information on each student's achievement

- be free from bias and provide evidence that accurately represents a student's knowledge, understanding and skills

- enable teachers and students to use feedback effectively and to reflect on the learning process

- be inclusive and accessible for all students

- be part of an ongoing process where progress is being monitored over time."

These principles form a foundation on which all good assessment should be based, including those formulating the on-call assessment element of the National Assessment Program. They are also consistent with what the second Gonski report, the aptly titled *Through Growth to Achievement*, envisaged as being in the best interests of school children: "Low stakes, low key and regular assessment … followed by teaching tailored to challenge the student to reach the next level of achievement."

The centrality of a shared, common curriculum as the basis for such assessments is also vital. A syllabus is a profound, broadly encompassing professional document that describes the depth and scope of an area of learning or discipline. It is the resource from which high-order programing can develop. It is also timely to note that syllabuses in NSW, for example, align with the Australian Curriculum.

The framework for on-call assessment

The proposal herein for the on-call component of the National Assessment Program would have the following characteristics:

- Schools in particular years would be required to undertake externally referenced assessments chosen from a section of the teaching program students have undertaken within their classroom in that year.

- The school would identify what is going to be assessed and when it is to be assessed.

- Schools would choose, from a national item bank, the assessments that they require.

- This item bank would be nationally developed and align with the Australian Curriculum.

- These item banks would need to allow for schools to request the specific, diverse and culturally appropriate assessment profiles in assessment items to correspond to the needs of students within a particular school and the teaching programs and differentiation that have been developed to respond to these needs and capacities. The range of these items will need to be tailored and finite to allow for their utility as assessment devices.

- These assessments should be capable of being conducted online, with pen and paper, or in oral or visual form.

- Over time, these assessment items should be developed to cover curriculum areas beyond literacy and numeracy.

- The bank of assessment items should be developed with the full participation of the teaching profession and should be controlled and disseminated by responsible public agencies.

- Marking of these tasks would be conducted by the teachers within the particular school or in a collaborative form with colleagues in other schools. Marking would be governed by the externally referenced marking criteria and work samples supplied with the test items.

Extensive professional learning would be provided to teachers to support this marking.

Assessing science: A case study

It is worth pausing at this point to examine what the current NSW practice in system-wide assessment of Science has provided in terms of an evidence base for future policy consideration. Since the current NAPLAN only focuses on literacy and numeracy, those two learning areas have dominated thinking but we have contemporary practice in Science assessment that is clearly superior in terms of practice and outcomes.

In NSW, statewide assessment of Science features the Validation of Assessment for Learning and Individual Development (VALID) program across the primary and junior secondary years. The title of the assessment gives a useful indicator of its different policy focus.

To quote from the NSW Department of Education's brief on VALID:

> "VALID delivers three online tests. VALID Science and Technology 6, VALID Science 8 and VALID Science 10. [The numbers correspond to the year the assessments are conducted with.]

> The VALID tests allow schools to map their students' progress in the Science [key learning area] against the assessment framework that spans Stages 2 to 5. The assessment framework describes six levels of achievement based on the educational theorem known as the Structure of Observed Learning Outcomes (SOLO).

> Assessment items are framed in real-life situations related to self, the family and the community. The items give students the opportunity to demonstrate their understanding at both local and global levels.

> The test also includes survey questions which assess the values and attitude outcomes of the science syllabus. All three VALID tests are interactive, online, multimedia diagnostic assessments. The tests contain both:

> • Extended response tasks, which provide an opportunity to assess higher order thinking and deeper understanding of a scientific concept or big idea. Secondary students attempt three extended response tasks. Primary students attempt two tasks.

> • Short response and multiple choice items presented in graphics, audio and/or text. All items in a set are contextually linked to the stimulus … Students are able to choose their own pathway through the item sets.

> VALID Science and Technology 6 and VALID Science 10 involve in-school marking of the extended response tasks … Teachers at schools participating in VALID Science and Technology 6 and VALID Science 10 need to undertake NESA professional learning in the differentiation of the quality of student responses.

> Following the professional learning, teachers analyse and mark the student responses for their school … At the completion of the in-school marking operation, schools will receive preliminary results prior to the release of the full data package on the School Measurement, Assessment and Reporting Toolkit (SMART) or Scout … which provides teachers with quality feedback on students' knowledge, understanding and skills, and identifies possible areas for future teaching and learning. Survey results reported at school and state levels are also provided through SMART/Scout."

It is also of note that the VALID tests are developed in consultation with all education systems in NSW; the teaching profession, students, parents and community along with academic, pedagogy, curriculum and assessment specialists.

The VALID assessments have been a significant success. They have developed quietly without the controversies and problems that have beset NAPLAN. They have utilised teacher expertise and have a successful school-based marking component that has contributed to high quality, school-based and statewide evidence collection.

The program has avoided the regressive effects of reporting on My School while still enabling parents, teachers and policymakers to track the results for individuals, school communities and identified equity groups. VALID has an on-demand component and a required level.

To understand more deeply the approaches that VALID has deployed, it is worth dwelling on a very recent study of Science assessment conducted by James Scott of the University of Technology Sydney (UTS) in his doctoral thesis on the subject.[6]

Scott's work spent time concentrating on Science assessment including, "a mandatory low stakes, large-scale, test-based diagnostic assessment program involving Year 8 … The assessment framework was fully implemented

across NSW from 2007". Note here the combination of "low stakes, large scale".

Scott reports: "The findings show that teacher use of three of five dimensions of formative practice and an explicit focus on teaching students the skills of writing to learn science produced science test results that were above expectation."

Scott found that there was a great opportunity now to build on the VALID approach using existing resources and technology:

> "The capacity to upload a wider range of responses to items and tasks would be made easier by transforming the once-a-year test to an online repository of items, related stimulus materials and extended response tasks from which teachers could choose. They could retain and store items online until they enabled access for their students as they work through the topic or at the end or both. The capacity for immediate feedback on their learning, this being one of the most powerful means for supporting learning, would then be provided. There are already a number of items (and related stimulus materials) and extended (open-ended) response tasks going back to 2005 held by the NSW Education Standards Authority that could be used to populate such a repository."

Scott further outlines the advantages of this online repository that would provide the engine for on-call assessments that would include the capacity to:

- provide immediate feedback to teachers about student experience of science

- provide a brief description of item and task lines to curriculum intentions

- inform about the level of cognitive demand of the item or task and possible real-world situations where engaging with the particular item and its stimulus material or task has benefits for the individual, society or the environment

- provide explanations of alternative conceptions indicated by student selection

of particular distractors in feedback to students and teachers

- facilitate suggestions for activities to correct misconceptions

- provide a range of answers that would be scored at different levels according to the solo model

- utilise the history of item and task use and student answers online and make it accessible to both teachers and the education system for monitoring purposes.

All of these benefits would apply equally to literacy and numeracy testing.

Some further findings made by Scott are of particular relevance to the purposes and Terms of Reference of the current review.

1. At the conclusion of a VALID assessment, "Parents (and their students) received a progress report about the learning in terms of both syllabus expectations and levels of understanding demonstrated in relation to those expectations … Teachers received a comprehensive analysis of individual performance on every task and item in the test as well as students' collective views about Science and their experience of it at school. Teachers were expected to use the results of the test and the survey to diagnose strengths, weaknesses and gaps in student learning (and level of engagement with learning science) and to respond accordingly."

2. "VALID data is provided to NSW schools and not published in the same way as NAPLAN data (*on a school-specific website for all the world to access* [his emphasis]). The findings reported in chapter five were that Science teachers understood the purpose of the VALID test, were willing to engage with it and feedback from it and appreciated the absence of pressures experienced by their colleagues more directly associated with the publication of NAPLAN results."

3. "In the event that public accountability is seen as important, consideration could be given to sample testing."

There is much that can be learnt from the clear evidence base provided by the achievements of the VALID Science Assessment program in NSW.

Global best practice

In addition to what we can learn from outstanding assessment practice in Australia, if we cast our gaze overseas we can see where the wisest policies are unfolding.

Singapore

Fiona Mueller, director of the education program at the Centre for Independent Studies, counsels us all to emulate Singapore, for example. As she advised us in The Australian newspaper:

"Singapore's path is one that should inspire Australia, and today it is clear that we need to make changes that will restore public confidence in our school education."[7]

We could not agree more.

In recent times, Singapore[8] has:

• eliminated all assessments and exams for the early primary school years

• reduced exams in secondary schools

• reduced the number of summative, in-class assessments

• increased the use of low stakes in-class assessments

• eliminated ranking students in relation to class or cohort

• established policy so that subject and group averages, overall total marks and minimum and maximum grades are set to disappear

• sought to, "allow each student to focus on his or her learning progress and discourage them from being overly concerned with comparisons".

As Ong Ye Kung, Singapore's Education Minister, puts it: "Learning is not a competition." The reforms in Singapore he describes as "another step towards improving the balance between the joy of learning and education standards".

"The goal of the changes is to give teachers more space to explore new ways of making learning enjoyable and lasting."

As for teachers, Ong Ye Kung observes:

> "On the ground, teachers are on a high speed train ... rushing assessments and preparing the students for exams ... I think it is a time for a pause."[9]

As the Singapore Ministry of Education calculated, these reforms will free up about three weeks of curriculum time every two years.

The same opportunity awaits Australia.

New Zealand

Our Australasian cousin provides further illustration of how other nations are adopting more sophisticated approaches to assessment.[10]

Choosing to move substantially away from the mass testing phenomenon, New Zealand is now placing far greater emphasis on supporting overall teacher professional judgement as a part of its assessment reforms.

Commencing with a strong foundation of professional learning support and ongoing resource availability, New Zealand is taking a systematic approach to the processes of gathering, analysing, interpreting and using information about students' progress and achievement. Teachers are provided with guidance around ways to increase the dependability of judgements including an emphasis on using a variety of sources in

arriving at deep understanding of student growth as well as insisting that the new approaches "retain at least the same rigour and educative purpose as before".

In striking this new flexible, evidence-based policy, teachers are advised quite clearly that:

• "Making judgements involves both student and teacher.

• Evidence is collected cumulatively over the year, in contexts across the curriculum and is brought together to judge achievement in relation to the New Zealand Curriculum.

• There is a need for information from a range of assessment approaches so that decisions are dependable.

• An overall teacher judgement is used to determine where a student's achievement sits in relation to the New Zealand Curriculum.

• Moderation improves both the dependability of [overall teacher judgements] and the evidence that supports them.

• Teacher curriculum and pedagogical content is essential for making a dependable [overall teacher judgement].

• Overall teacher judgements, constructed with students, are the basis of the reporting in relation to the New Zealand Curriculum."[11]

> **For further information about developments in New Zealand see Appendix B**

Scotland

As a case study of how low-stakes, comprehensive, on-demand assessments can enhance educational standards, Scotland now provides a valid and reliable precedent.

In order to provide the externally referenced component of what Dylan Wiliam regarded as essential to a humane but rigorous assessment regime, the development of the

Scottish National Standardised Assessments offers inspiration.

Under Scottish National Standardised Assessments, children in years that approximately correspond to our NAPLAN test years undertake these assessments of literacy, writing and numeracy.

The assessments are:

• purely diagnostic

• removed from concepts of pass or fail

• marked online with rapid feedback provided to teachers

• aligned entirely with the relevant Scottish curriculum

• adaptive

• conducted at the time the school regards as most useful

• designed so that no extra workload for teachers or students is required

• of short duration to avoid unnecessary stress

• constructed "so daily classroom learning is all the preparation a child or young person will need"

• supported by professional learning programs for teachers

• protected from prurient media reporting and league tables.

The Scottish government states: "The assessments help to identify children's progress, providing diagnostic information to support teachers' professional judgement … to provide teachers with objective, comparable information about progress, which will help improve outcomes for all children and young people."

The assessments are made available to schools between August and the end of the year. The tests are quickly marked and the information returned on individuals or groups of students.

Most happily, this assessment system is provided to Scotland by the Australian Council for Educational Research so the intellectual capital and personnel is on hand in this country.

Nicola Sturgeon, Scotland's First Minister, sees these reforms as central to her goal of improving Scottish educational outcomes, which she describes as "arguably the most important objective" for her government.

As her administration explains:

> "These assessments provide teachers, for the first time, objective and nationally consistent information on children's progress in aspects of literacy and numeracy, alongside a wide range of other assessment activity. Teachers can then discuss children's progress with them and their parents, taking into account the full range of assessment activity including [Scottish National Standardised Assessments], to plan next steps and ensuring parents understand how best to support their child's learning at home."[12]

Agreeably, these international best-practice directions are consistent with what we have learnt through the process surrounding the second Gonski report, *Through Growth to Achievement*.

Its Finding 7 dovetails elegantly with what is evolving overseas: "There is compelling evidence, in Australian schools and internationally, that tailored teaching based on ongoing formative assessment and feedback is the key to enabling students to progress to higher levels of achievement." While its Recommendation 11 calls for the development of "a new online and on-demand student learning assessment tool … [supporting] low stakes, low key and regular assessment followed by teaching tailored to

challenge the student to reach the next level of achievement".

> *For further background to developments in Scotland see Appendix C*

Student-centred reporting

Naturally, we cannot develop sound assessment policy without regard to its essential concomitant, a purposeful, humane and student-centred system of reporting.

It is the clear position of this submission that **parents have an absolute right to know** how their own child is progressing and what should be done in order to advance their child's development and intellectual advancement.

It is the **responsibility of teachers** to engage in regular, purposeful assessment and testing in order to ascertain how their students' progress is unfolding, how effective their teaching programs are, and to provide detailed and meaningful reports to parents.

It is the **responsibility of school systems, public officials and public authorities** to track how their schools are working, how particular communities are faring, what trends are emerging, how particular equity groups are performing and consequently where extra support and resourcing should be directed.

The current, unhappy state of affairs

Currently, we have in many parts of the nation a common, reductive, hierarchical and regressive format known as A-E reporting.

Under this dictate, students are placed on a continuum wherein they are often locked into a rigid, archaic mode of ranking. Within this system, a battling primary school student might have been given more than a hundred dispiriting low grades by the time they enter secondary school and all under the school's crest and over the signature of the principal.

In its submission to the Gonski 2.0 investigations, the Australian Council for Educational Research made its position clear in relation to the effects of such reporting:

> "The problem with A-E grades and similar methods of reporting is that they do not show where students are in their long-term learning or progress over time. A student can receive the same grade (e.g. a grade of D) year after year. The first problem with this is that it does not enable students to see the absolute progress they are making (and in fact hides this progress). The second problem is that this method of reporting risks sending a message that there is something stable about the student's ability to learn (e.g. they are a D student)."

Elsewhere, Australian Council for Educational Research also argues that A-E reporting sends the worst of messages to a vast range of students transmitting counterproductive signals about effort and reward. The battling student can gain a profound sense of sustained failure often leading to disappointment, a sense of futility in their learning and deepening disengagement. Meanwhile, very able students may consistently gain high grades based on minimal effort and draw the inference that they need not work hard at their own personal learning and they can develop a nonchalant attitude to their future learning, often with regrettable outcomes.

Along with the National Assessment Program featuring a paired system of mass sample testing and on-call, school-based assessment, it is now imperative that we have a reporting process that eschews the current counterproductive regimen of public shaming and private bewilderment.

Instead, reports must now focus on growth and individual achievement.

Protocols of reporting

Alongside the reforms suggested in this submission must come a new understanding of who "owns" the results of a worthwhile assessment system. Certainly, policy makers must make themselves aware of how their schools are performing and where improvement and further support must be directed.

For parents and carers the following principles should apply to the results of the assessments their children have undertaken:

1. Parents have an absolute right to know how their child is progressing.

2. Parents have an absolute right to know how their child can be supported at school and at home to further their achievement.

3. Parents have an on-request right to see all detailed results of their child's assessments, including school-based and externally referenced assessments.

4. Parents do not have a right to see the results of other children's or schools' performances.

5. The public has a right to know that teachers, schools and responsible authorities are collecting and acting upon the results of assessment programs.

6. Children have the right to have their private academic performance protected from public gaze along with other rights to data privacy we all share as citizens.

7. When the state mandates that a child must undertake a test then the state has a profound moral obligation to protect the child from breaches of confidentiality surrounding those private academic assessments.

The reckless culture of invading the privacy of children that the current NAPLAN regime has engendered is in fact in stark contrast to the recent, proud heritage of test data protection that used to prevail in NSW, for example.

With the introduction of Basic Skills Testing in that state in the 1990s, there were stringent and bipartisan statutory protections of the results of these tests.

The Coalition education minister at the time, Virginia Chadwick, insisted on a regulation within the Education Reform Act 1990 relating to these tests. The regulations were clear and had child protection at their core, as this section of the Act indicates:

"Basic Skills Testing

6. The following provisions apply with respect to the confidentiality of the results of basic skills testing under section 18 of the Act:

(a) The results of a particular child may be revealed only to the child, to the child's parents and to the school in which the child is currently enrolled;

(b) The results of individual schools must not be publicly revealed;

(c) Aggregate statewide results may be publicly revealed if comparisons are not made (and not capable of being made) between different children, different schools or different systems or groups of schools."

When an ALP government came to power somewhat later they enhanced the provision.

The education minister, John Aquilina, introduced a further regulation to the Act in 1996:

"5. (3) the results of basic skills testing must not be publicly revealed in a way that ranks or otherwise compares the results of particular schools."

In the following year, Aquilina extended these protections to the results of HSC students and these principles are still in place today for HSC students.

But not so for children as young as eight years old, who undergo mandatory testing and then have these results cast out into the public domain.

As a result of coercive federalism within the national schools funding agreements and the coming of the NAPLAN/My School monolith, jurisdictions have been compelled to remove their statutory protections for children's formerly private test outcomes.

In the decades since children were first provided with these legal safeguards, concerns about data collection, privacy, pervasive technology, irresponsible journalism and social division have increased mightily.

It is instructive to note here that tabloid newspapers continue to campaign to have even more intrusive powers to report the academic results of young children. *The Daily Telegraph* recently lamented that it could not publish league tables in NSW like News Limited newspapers in other jurisdictions.

The Daily Telegraph[13] attacked what it described as "repressive legislation" that did not allow them to sufficiently disclose NAPLAN and HSC results and constrained them from publicly ranking schools. Citing a spokesperson for the Institute of Public Affairs, the tabloid reckoned that not publishing such league tables "let teachers dodge any criticism of what goes on in the classroom".

The newspaper was clear in its campaign goal: "It is time that the law was changed." This News Limited campaign manifested itself again across Australia in their various publications in December, 2019.

The time is surely overdue that we reintroduced these principles of legal protection for children and reinstated in appropriate legislation the regulatory confinement of student test data.

We would never allow this invasion of privacy of adult citizens that currently is annually inflicted on young children and their communities in the name of NAPLAN and My School.

My School

In addition to redirecting school-based reporting to focus on the growth, needs and achievement of the individual child it is imperative to reform the public spectacle that reporting My School has become.

It is useful to remind ourselves of its provenance. My School was the brainchild of then-federal education minister Julia Gillard and the US education pundits who were in vogue for a regrettable moment. The major purpose of the website, as revealed in Gillard's second reading speech in 2008, was to drive school choice. This created Australia as one educational market and destroyed NAPLAN as a diagnostic tool overnight.

It was another market theory fix attached to the school lives of children. National public exposure of school results was supposed to drive school performance, and parents as consumers would use the website as a catalogue to flick through before landing on a school product that pleased them.

As Rob Stokes observed when in office as NSW education minister: "My School is now used dishonestly as a school rating system."

School students and communities have now endured a decade of this theory. By the time many students have now completed year 9 they have been told repeatedly and publicly and in technicolour that they are no good at learning. Many have heeded the lesson.

It is timely to note the conclusions drawn by Australian Council for Education Research chief executive officer Geoff Masters on what the decade of NAPLAN/My School has wrought.

Professor Masters observed in an analysis in *The Conversation* (May 16, 2018):

> "Since the introduction of NAPLAN, there has been a marked increase in the stakes attached to these tests. School results have been made available for public comparison on the My School website. Some schools even use NAPLAN in their marketing and school selection processes.

> "Parents, teachers and schools now place greater emphasis on NAPLAN results in comparison to earlier state-based tests. This has led to reports of inappropriate levels of practice testing and increased student test anxiety. It has also narrowed teaching to the test, and led to occasional cheating.

> "The decision to make all schools' NAPLAN results public was based on a belief this would provide parents with better information when choosing schools.

> "This is a market-driven belief that, for schools, the risk of losing students would be a powerful incentive to improve. But test-based incentives have proven largely ineffective in driving school improvement.

> "Parents have sometimes drawn incorrect conclusions about the quality of a school from publicly reported test results. And public comparisons of schools have resulted in a range of unanticipated negative consequences such as narrowing teaching and increasing levels of teacher and student stress.

> "An obvious strategy is to stop reporting school results publicly and to restrict access to school-level data to individual schools and school systems. The primary focus of literacy and numeracy testing might then return to its original purpose of informing teaching and learning."

Professor Masters has clearly established that My School has failed in one of its original stated purposes of driving enhanced school performance.

It has succeeded, however, in its application of market theory to education. Markets so misapplied will always generate winners and losers. In this context, some of the losers are as young as eight.

A visitor to the shopfront of the market, the My School website, can wander through the private academic performance of groups of children who are forced to undertake a state-mandated test. Anyone can now idle away hours hypothesising how this or that school community performed in their tests and then to speculate, often quite mindlessly, as to what

type of school it must be and what conclusions might be.

With three clicks anyone can discover the test outcomes, the level of disadvantage, the cultural backgrounds, the language backgrounds, and the level of indigeneity of the children within any school community.

One could scarcely design a more efficient tool for social division and increased segregation and snobbery. And remember, of course, that it was deliberately designed as a device for school choice. And it is colour-coded just in case one misses the point.

In Australia, this is especially reprehensible at a time when public policy should be working towards social cohesion and providing hope and encouragement for our young, in particular, with a world ever-more fractured and where mindless populism threatens to divide whole nations.

And there is indeed emerging evidence that My School is contributing to a more divided Australia. As the analysis of NAPLAN/My School provided by Chris Bonnor reveals:

• Equity slopes are increasing (worsening) for almost every NAPLAN domain for every main location and every sector.

• High ICSEA schools are accumulating even more high ICSEA quarter enrolments (the most advantaged students) while low ICSEA schools are losing them and gaining more lowest quarter enrolments (up from 54 per cent to 72 per cent).

• Rural and remote schools are accumulating more low ICSEA students.

• Government schools are getting a higher proportion of low ICSEA students and a lower proportion of high ICSEA enrolments.

Bonnor's work demonstrates that My School is contributing significantly to a more divided, socially segregated Australia.

The grim irony here is that drawing any conclusions based on the NAPLAN test is fraught. They are often low grade tests with low standard marking criteria as established in respect of the readily gamed writing component by Les Perelman.[14]

Or, as Margaret Wu[15] has found, the results themselves can be wildly inaccurate and varying participation rates can render any conclusions drawn extremely doubtful.[16]

The devil in the clicks can be revealed by those persistent enough to submerge themselves in the technical detail held deep inside My School. One can uncover a strangely low confidence interval attached to the results. A 90 per cent confidence level is used across all NAPLAN data on the My School website, which is a much lower level of confidence than the standard 95 per cent confidence interval used in most social and educational research.

Consequently, as Nicole Mockler proves, the 90 per cent confidence interval covers approximately two bands of NAPLAN, scores that would change the colour palette of the results entirely and render any educational or dinner party conclusions completely invalid. Mockler uses an example from the 2016 NAPLAN technical report whereby a score of 615 in grammar and punctuation is in fact a score within the range of 558 and 672 at the 90 per cent confidence level.[17]

Even by the key performance indicators that My School is purported to achieve, it fails itself, and the Australian student population, woefully.

One will also note that there is no mention herein of the validity of results in high-order capacities — creativity, imagination, engagement, critical thinking, synthesis of knowledge — attributes that we should be seeking to value and nurture in all of our children. NAPLAN and hence, My School, never travels there.

The current state of affairs persists with the unique and perverse quality of NAPLAN/My School being both low grade and high stakes.

Unsurprisingly, after the failed decade of NAPLAN/My School, the teaching profession across all sectors is deeply concerned about the negative role that My School plays in school communities. Surveys of teachers conducted by the Australian Education Union now regularly show that more than 80 per cent of teachers believe that the test data of their students should not be made publicly available.

The same level of concern exists among private schools where a national 2019 survey of heads of schools revealed that more than two thirds of them believed that NAPLAN data should not be published on the My School website.

As The Sydney Morning Herald reported:

> "68 per cent said the data should not be published on My School. They were most worried that it conveyed a narrow understanding of a successful education, a narrow view of student achievement, and a short-term view of student progress."[18]

So, what should be done?

The following framework, to be read in conjunction with the Protocols of Reporting, is proposed to replace the existing My School.

A new My School

1. Students' right to academic privacy should be protected in law.

2. My School can be used by parents to gain an overview, provided by each school, of the philosophies, emphases and approaches used by each school in the interests of their students.

3. The best way to find out about a school is to visit that school and have detailed discussions with teachers.

4. The My School website will provide contact details for each school in Australia and provide the means to make an appointment with that school to have further discussions.

5. Data that is currently displayed on the My School website will continue to be collected by responsible public authorities to inform policy considerations.

6. Such information will also be available, on a confidential basis, to academics, public policy specialists and researchers working in the interests of education and students in Australia.

Conclusions and new directions

A basis for a new National Assessment Program framework

It is proposed that a new National Assessment Program requires the following elements:

1. A system of national assessment, regularly conducted, based on a huge, comprehensive and inclusive sample of students deploying a full range of levels of intellectual challenge that provides parents, teachers and public officials with a clear understanding of how various social groups, jurisdictions, and parts of the country are progressing. Initially these assessments would concentrate on literacy and numeracy but would expand over time to assess other parts of the curriculum.

This would represent Australia's Report Card.

2. All schools would be additionally required to regularly assess how their students are progressing by undertaking on-call assessments, supplied from a national bank of test items that are aligned to the Australian Curriculum. This on-call testing would be on focus areas determined by the respective schools and teachers and based on the specific program of teaching experienced in classrooms. It is expected that literacy and numeracy would be given prominence in these on-demand assessments but other learning areas would also be developed. The results of these tests would be individually available to parents and caregivers of students and give a picture of how well a child was progressing in areas actually taught in their classroom and also provide a progress report of how a child was growing in relation to the national standards.

This component relies on the simple but profound principle of *testing to the teaching* rather than the current unhappy state where so many schools feel compelled to teach to the test.

The essence of the paired testing system is to provide valid and reliable evidence of both how the nation is advancing and how each individual child is progressing.

Australia can learn a great deal in the development of these paired testing processes from the best, cognate, international practice and this paper will outline these global exemplars. This paper will outline significant intellectual precedent that provides the research evidence base for all of the proposals herein, gathered from Singapore, the US, Scotland and New Zealand.

One can infer from the evidence of existing practice that moving to low impact, sample testing would add on average one week's teaching time for every student in the land.

An additional boon would be that the many millions of dollars that are spent across Australia to deliver the current NAPLAN census test regime could be redistributed to fund the improvements in assessment that this Review now has the chance to create.

The framework for on-call assessment

The proposal herein for the on-call component of the National Assessment Program would have the following characteristics:

- Schools in identified years would be required to undertake externally referenced assessments chosen from a section of the teaching program students have undertaken within their classroom in that year.

- The school would indicate what is going to be assessed and when it is to be assessed.

- Schools would choose, from a national item bank, the assessments that they require.

- This item bank would be nationally developed and align with the Australian Curriculum.

- These item banks would need to allow for schools to request the specific, diverse and culturally appropriate assessment profiles in assessment items to correspond to the needs of students within a particular school and the teaching programs and differentiation that have been developed to respond to these needs and capacities. The range of these items will need to be tailored and finite to allow for their utility as assessment devices.

- These assessments should be capable of being conducted online, with pen and paper, or in oral or visual form.

- Over time, these assessment items should be developed to cover curriculum areas beyond literacy and numeracy.

- The bank of assessment items should be developed with the full participation of the teaching profession and should be controlled and disseminated by responsible public agencies.

- Marking of these tasks would be conducted by the teachers within the particular school or in a collaborative form with colleagues in other schools. Marking would be governed by the externally referenced marking criteria and work samples supplied with the test items. Extensive professional learning would be provided to teachers to support this marking.

Science lessons

There is much that can be learnt from the clear evidence base provided by the success of the VALID Science Assessment program in NSW.

The VALID assessments have been a significant success. They have developed quietly without the controversies and problems that have beset NAPLAN. They have utilised teacher expertise and have a successful school-based marking component that has contributed to high quality, school-based and statewide evidence collection.

The program has avoided the regressive effects of reporting on My School while still enabling parents, teachers and policymakers to track the results for individuals, school communities and identified equity groups. VALID has an on demand component and a required level.

Student-centred reporting

Naturally, we cannot develop sound assessment policy without regard to its essential concomitant, a purposeful, humane and student-centred system of reporting.

It is the clear position of this submission that **parents have an absolute right to know** how their own child is progressing and what should be done in order to advance their child's development and intellectual advancement.

It is the **responsibility of teachers** to engage in regular, purposeful assessment and testing in order to ascertain how their students' progress is unfolding, how effective their teaching programs are, and to provide detailed and meaningful reports to parents.

It is the **responsibility of school systems, public officials and public authorities** to track how their schools are working, how particular communities are faring, what trends are emerging, how particular equity groups are performing and consequently where extra support and resourcing should be directed.

Protocols of reporting

Alongside the reforms suggested in this submission must come a new understanding of who "owns" the results of a worthwhile assessment system. Certainly, policy makers must make themselves aware of how their schools are performing and where improvement and further support must be directed.

For parents and carers the following principles should apply to the results of the assessments their children have undertaken:

1. Parents have an absolute right to know how their child is progressing.

2. Parents have an absolute right to know how their child can be supported at school and at home to further their achievement.

3. Parents have an on-request right to see all detailed results of their child's assessments, including school-based and externally-referenced assessments.

4. Parents do not have a right to see the results of other children's or schools' performances.

5. The public has a right to know that teachers, schools and responsible authorities are collecting and acting upon the results of assessment programs.

6. Children have the right to have their private academic performance protected from public gaze along with other rights to privacy we all share as citizens.

7. When the state mandates that a child must undertake a test then the state has a profound obligation to protect the child from breaches of confidentiality surrounding those private academic assessments.

A new My School

1. Students' right to academic privacy should be protected in law.

2. My School can be used by parents to gain an overview, provided by each school, of what the philosophies, emphases and approaches used by each school in the interests.

3. The best way to find out about a school is to visit that school and have detailed discussions with teachers.

4. The My School website will provide contact details for each school in Australia and provide the means to make an appointment with that school to have further discussions.

5. Data that is currently displayed on the My School website will continue to be collected by responsible public authorities to inform policy considerations.

6. Such information will also be available, on a confidential basis, to academics, public policy specialists and researchers working in the interests of education and students in Australia.

The time is surely overdue that we reintroduced principles of legal protection for children and reinstated in appropriate legislation the regulatory confinement of student test data.

We would never allow this invasion of privacy of adult citizens that currently is annually inflicted on young children and their communities in the name of NAPLAN and My School.

Endnotes

1. See Ainley and Gebhardt Measure for Measure: A review of outcomes of school education in Australia 2013 and Masters "Five Challenges in Australian School Education" 2016

2. "NAPLAN a serious distraction: principal" *The Sun Herald* 9 September, 2018, p10

3. Tim Dodd "After 10 years it's time to ask what is the purpose of NAPLAN? *Australian Financial Review 30* July, 2017

4. Hilary Persky "Writing Assessment in the Context of the National Assessment of Educational Progress" Educational Testing Service undated paper

5. Dylan Wiliam "The bridge between teaching and learning" in Pedagogiska Magasinet Sweden 16 September, 2011

6. "Exploring the impact of a large-scale diagnostic science test and formative practices: a mixed methods study" *https://opus.lib.uts.edu.au/handle/10453/133155*

7. *The Australian* 31 August–1 September, 2019

8. See *straitstimes.com/singapore/education/5-key-changes-to-examinations-and-assessments-in-schools* Accessed 15 October, 2018

9. See *straitstimes.com/singapore/education/fewer-exams-for-students-less-emphasis-on-grades* Accessed 15 October, 2018

10. See *assessment.tki.org.nz*

11. See *assessment.tki.org.nz/Overall-teacher-judgment/Increasing-dependability* Accessed 12 August, 2019

12. See *standardisedassessment.gov.scot/* Accessed 29 July, 2019

13. "NAPLAN ban fails state test" *Daily Telegraph* 5 October, 2019 p17

14. See Dr Les Perelman Towards a New NAPLAN:Testing to the Teaching NSW Teachers Federation 2018

15. See "Interpreting NAPLAN results for the layperson" *edmeasurement.com.au/_publications/margaret/NAPLAN_for_lay_person.pdf* Accessed 12 March, 2019

16. See G Thompson, L Adie and V Klenowski "Validity and participation: implications for school comparisons of Australia's National assessment Program: Journal of Education Policy 33 (6) 2018 pp 759-777

17. N Mockler "It's time to be honest with parents about NAPLAN; your child's report is \misleading, here's how" *AARE aare.edu.au/blog/?p=2870*

18. Jordan Baker "Majority of private school heads oppose publishing NAPLAN" *Sydney Morning Herald* 27 March, 2019
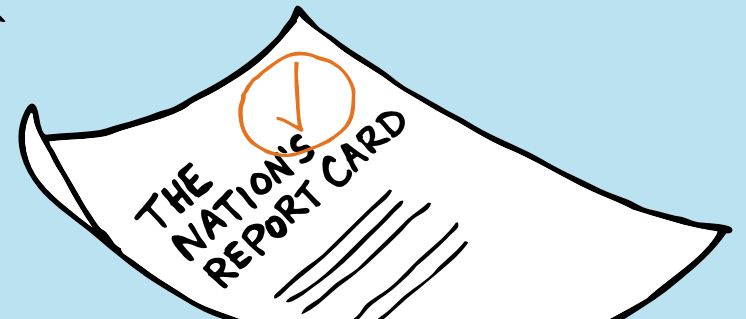
Towards a Better National Assessment Program

Appendices

Appendix A

The Nation's Report Card

What Every Parent Should Know About NAEP
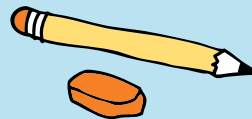
You've probably heard of The Nation's Report Card.

It's a resource. It's a national wakeup call. It's a window into the state of our educational system and what students are learning.

It's more than just a test.

Today's students take tests for everything. **The National Assessment of Educational Progress (NAEP)**, though, is different. It's called The Nation's Report Card because it tells us how students across the country, not just in one particular school or state, are doing.
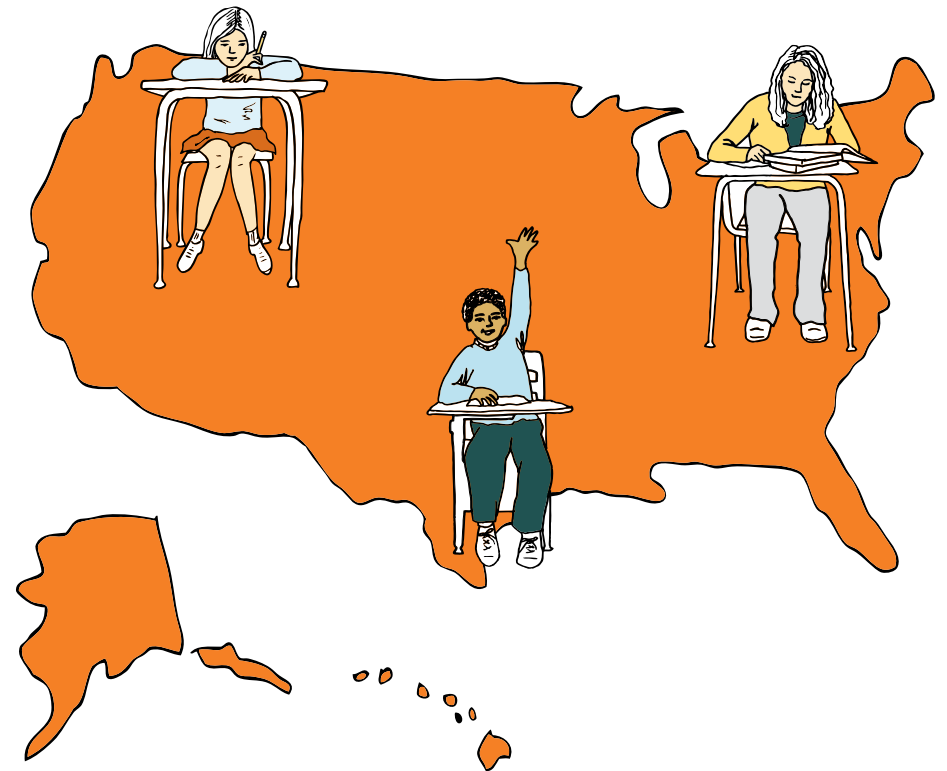
"[T]he definitive source... for state-to-state comparisons in educational achievement..."

EDUCATION WEEK

-Influence: A Study of the Factors of Educational Achievement

"Only 12% of U.S. high school seniors are proficient in the legacy of America's past."

abc NEWS

"...higher math scores are "undoubtedly" **a sign of the success of math instruction** in classrooms..."

THE WALL STREET JOURNAL.

"...urban school districts have steadily increased their scores since 2003..."

The Miami Herald

Everyone's talking about **NAEP**. As the largest national test, its results are found in the headlines of national and local news stories all over the country.

"**NAEP is the gold standard of student assessments.**"

The Washington Post

-The Washington Post editorial board

"Students taking the math exam who reported taking algebra I in eighth grade tended to outscore their peers who had not taken algebra I."

MILWAUKEE·WISCONSIN
JOURNAL SENTINEL

"The nation's report card on math and reading shows **fourth- and eighth-graders scoring their best ever in math**..."

ASBURY PARK PRESS

"Today's NAEP results confirm that **we have a crisis on our hands** when it comes to civics education."

The New York Times

"**Very few students have the advanced skills that could lead to careers in science and technology, according to the results of a national exam...**"

LAWRENCE JOURNAL-WORLD.

NAEP: The Big Picture

The Nation's Report Card opens the door for parents to understand how well children like theirs are doing compared to other students in:

- Participating large urban districts
- Other states
- The nation
- The rest of the world!

Your State!

MD

the world

The Nation's Report Card

participating large urban districts

the nation

Other states

TX

KS

CA

IL

NAEP in a Nutshell

- It is often called the gold standard of tests and has been around since 1969.

- The subjects most frequently tested are math, reading, science, and writing.

- Not every student or every school takes the test.

- The results are anonymous and if your child is asked to take it, his or her grades won't be affected.

- Accommodations are provided so that as many students as possible can participate.

How would my child stack up?
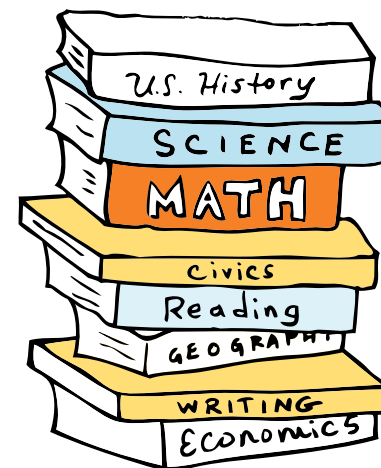
Use real NAEP questions to find out.

After each test, dozens of NAEP sample questions are released with The Nation's Report Card to show how real students performed in that subject.
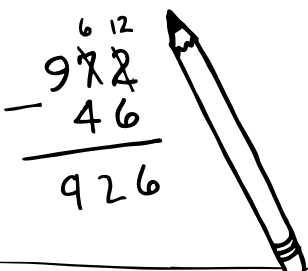
U.S. History

As part of the U.S. history assessment, students were asked the purpose of the Declaration of Independence.

34 percent of fourth-grade students answered correctly.

Mathematics

As part of the math test, students were asked to subtract 46 from 972 and fill in their answer.

76 percent of fourth-grade students answered correctly.

Give it a try. Test yourself!

I tried my best. Mom said it was important!

Science

Lightning and thunder happen at the same time, but you see the lightning before you hear the thunder. Explain why this is so.

Do you know the answer? Find out online!

ARE YOU SMARTER THAN YOUR FOURTH-GRADER?
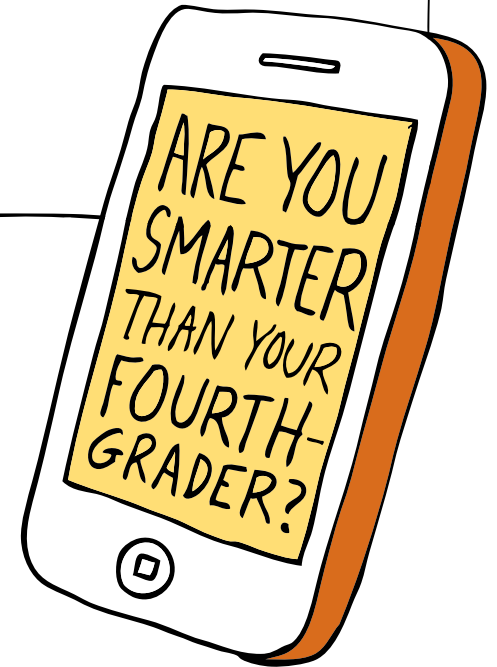
What do all of these NAEP results tell us about education?

See how your state measures up.

You can use NAEP results to compare your state to the nation, other states, and some large urban districts.
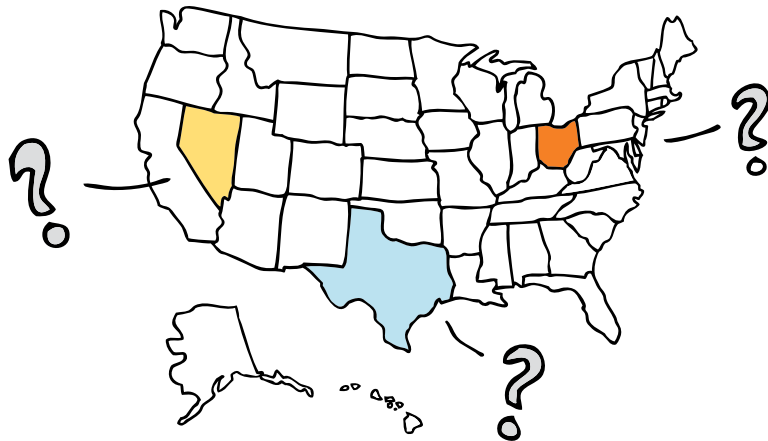
Go beyond the scores.

NAEP reports provide more than just results in school subjects and grades.

See how students have performed over time, including by racial/ethnic groups and gender.

NAEP also looks at differences between the performances of groups of students, illuminating achievement gaps that exist among the nation's students.

Learn which states are closing achievement gaps, and see whether your state is making progress.

NAEP examines the path to high school graduation by studying coursetaking and grades of America's most recent high school graduates.

Find out if your child is prepared for the future.

Look at the full picture.

NAEP also looks at the factors that may be related to learning—both inside and outside the classroom.

★ Fourth-graders who reported **reading for fun almost every day** scored higher on NAEP.

★ The percentage of twelfth-grade students who reported **studying the U.S. Constitution** in 2010 was lower than in 2006.

★ Eighth-graders who reported **taking algebra I** scored higher on average than students taking lower-level courses.

★ Since 1990, more high school graduates from each racial/ethnic group **completed a rigorous curriculum.**

★ 90 percent of fourth-graders who took the NAEP reading assessment in 2011 reported that they **had a home computer.**

★ More than any other single subject, **science was the key to reaching a higher curriculum level.**

NAEP Sparks Change Across the Country

Detroit, Michigan

In 2009, the NAEP assessment revealed that Detroit schoolchildren ranked the lowest in the nation in both grades 4 and 8. In response to the alarming results, *The Detroit Free Press* partnered with Detroit Public Schools (DPS) to create and implement a new reading initiative, The Call to Action for a new Reading Corps, which encouraged citizens to volunteer 100,000 hours collectively to tutor reading in DPS schools.

North Carolina

In 2005, results from the NAEP reading assessment revealed that eighth grade students in North Carolina scored below the national average. In response, the state deployed more than 200 literacy coaches to middle schools around the state to help teachers reach students with reading difficulties before they made the transition to high school.

Visit The Nation's Report Card online.
www.nces.ed.gov/nationsreportcard

Dig deeper into topics such as:

• How your child was selected
• NAEP accommodations
• NAEP scoring

Stay involved! Connect with NAEP online to be the first to know about upcoming assessments and reports.

www.ed.gov   ies.ed.gov

ies NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

NAEP

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

U.S. Department of Education
NCES 2013-455

National Assessment of Educational Progress

An Overview of NAEP

What is NAEP?

The National Assessment of Educational Progress (NAEP) is the largest nationally representative and continuing assessment of what students in the United States know and can do in various subjects. Since 1969, NAEP has provided a common measure of student achievement across the country. The results are released as The Nation's Report Card and are available for the nation, states, and in some cases, urban districts. NAEP is a congressionally mandated project administered by the National Center for Education Statistics (NCES), located within the U.S. Department of Education's Institute of Education Sciences.

How is NAEP different from state assessments?

NAEP serves a different role than state assessments. States have their own unique assessments which are designed to provide individual student data about achievement on different content standards. NCES administers the same NAEP assessment in every state. This provides a common measure of achievement that allows for comparisons of achievement to the nation and among states and participating urban districts. NAEP does not report results for individual students.

Unlike state assessments, students, teachers, and principals who participate in NAEP are also asked to complete contextual questionnaires. This information provides a better understanding of students' educational experiences and factors that may be related to their learning. To learn more about these questionnaires, visit http://nces.ed.gov/nationsreportcard/bgquest.aspx.

▸ The NCES NAEP website provides more extensive information about the assessment:
http://nces.ed.gov/nationsreportcard

▸ NAEP Tools on the Web provide quick and easy access to NAEP assessment data, previously administered NAEP questions, performance comparisons, and more:
http://nces.ed.gov/nationsreportcard/about/naeptools.asp

▸ Full copies of all reports are available on The Nation's Report Card website:
http://nationsreportcard.gov

What subjects does NAEP assess?

Main NAEP assessments are conducted in a range of subjects with fourth-, eighth-, and twelfth-grade students across the country. Mathematics and reading are assessed every two years, and science and writing are assessed every four years. Other subjects are assessed periodically, including the arts, civics, economics, geography, technology and engineering literacy, and U.S. history.

Long-term trend NAEP measures student performance in mathematics and reading, and is designed to ensure comparability across the years. The long-term trend assessment makes it possible to chart educational progress since the early 1970s. The assessment is given to 9-, 13-, and 17-year-olds every four years.

NAEP special studies, such as the High School Transcript Study, are conducted periodically in addition to main and long-term trend assessments. The National Assessment Governing Board, which sets NAEP policy, determines the assessment schedule and what content should be measured. To download a detailed PDF version of the assessment schedule, please visit http://www.nagb.org/newsroom/assessment-schedule.htm.

What can schools and students expect when participating in NAEP?

▸ NAEP is administered to students during regular school hours. Students spend between 90 and 120 minutes taking the assessment.

▸ Each student is assessed in one subject area and is asked to provide contextual information, such as the amount of reading they do, the types of classes they take, and their experiences with technology.

▸ Accommodations are provided for students with disabilities and/or English language learners.

▸ Student results are confidential. All materials are brought to and taken from the school by NAEP representatives on assessment day.

Why is student participation important?

The participation of all selected students enables NAEP to obtain an accurate picture of how students are performing in the United States. Since NAEP is not designed to report results for individual schools or students, it is not necessary for every student in every school to take the assessment. Instead, NCES uses a sampling procedure to ensure that NAEP participants are representative of the geographical, racial/ethnic, and socioeconomic diversity of schools and students across the country. Each participating student represents hundreds of other students. Their participation is vital for valid information on student achievement to be collected and shared. Teachers, principals, parents, policymakers, and researchers all use NAEP results to assess student progress and develop ways to improve education across the country.

How is NAEP using technology to measure and report the skills of a new generation of students?

As computers and digital tools play an increasingly important role in today's classroom, NAEP is advancing with technology-based assessments. The goal is to be paperless by the end of the decade. Through the following innovative assessments, NAEP is collecting new types of data that provide depth in our understanding of what students know and can do, including how they engage with technology to approach problem solving.

▸ **TECHNOLOGY AND ENGINEERING LITERACY (TEL) ASSESSMENT:** TEL is a new frontier for NAEP and large-scale assessment. It is a computer-based, cross-curricular assessment that challenges students to perform interactive tasks and engage in solving problems within realistic scenarios. TEL gauges how well students understand and apply technology and engineering principles to real-life situations. To learn more about TEL, visit http://nces.ed.gov/nationsreportcard/tel.

▸ **WRITING ASSESSMENT:** The writing assessment is administered on a computer and asks students to respond to writing prompts delivered in multimedia formats, including short videos and audio. In addition to writing scores, assessment results provide information about the extent to which students engaged in certain actions on the computer as they responded to these tasks, such as thesaurus tool usage. Results and contextual information are available at http://nces.ed.gov/nationsreportcard/writing.

▸ **SCIENCE INTERACTIVE COMPUTER TASKS (ICTs):** The NAEP science assessment includes ICTs that challenge students to solve scientific problems and perform experiments, often by simulation. ICTs provide students more opportunities than a paper-and-pencil assessment to demonstrate skills involved in doing science, without many of the logistical constraints associated with a natural or laboratory setting. The full library of released ICTs from the 2009 assessment is available at http://nationsreportcard.gov/science_2009/ict_tasks.asp.

After each assessment, some NAEP questions are released to the general public along with data on how students performed on each question. The following are examples of NAEP questions and actual correct responses from students.

4th Grade Reading
For the following question, students read a passage from the story "Tough as Daisy," which is about a young girl who has moved to a new school and must prove she is a good enough wrestler to be on the team.

At the beginning of the story, when some of the boys point and laugh at Daisy, she thinks, "We'll see about that." What does this tell you about Daisy?

What this tells me about Daisy is she is confident and strong. She never gives up. She never thinks she is bad at anything.

8th Grade Science
This multiple-choice question measures students' performance in the Earth and space sciences content area.

These funnels were filled with equal volumes of pebbles, fine sand, and coarse sand, as shown in the diagram below. The same amount of water was poured into each funnel.

Pebbles   Fine Sand   Coarse Sand

Which correctly lists the order in which the water passed through the funnels, from fastest to slowest?

Ⓐ Pebbles, fine sand, coarse sand   Ⓒ Fine sand, coarse sand, pebbles
Ⓑ Pebbles, coarse sand, fine sand   Ⓓ Coarse sand, pebbles, fine sand

12th Grade Mathematics
The following is a short constructed-response question from the algebra content area.

If $f(x) = x^2 + x$ and $g(x) = 2x + 7$, what is an expression for $f(g(x))$?

$$f(2x+7) = (2x+7)^2 + (2x+7)$$

An Introduction to **NAEP**

National Assessment of Educational Progress

"**NAEP is a necessary tool and way for my colleagues and me to make sure that the expectations that we have for teaching, learning, and ultimately student achievement, are where they belong.**"

Brian Dassler, Principal
KIPP Renaissance High School, New Orleans, LA

TABLE OF CONTENTS

The National Assessment of Educational Progress (NAEP) is a continuing and nationally representative assessment of what our nation's students know and can do. NAEP has often been called the "gold standard" of assessments because it is developed using the best thinking of assessment and content specialists, education experts, and teachers from around the nation.

NAEP provides a common measure of student achievement across the country. Because states have their own unique assessments with different content and standards, it is impossible to use them as a means for comparing state achievement. Such comparisons are possible with NAEP, however, because the questions and administration of the assessment are the same across all states.

NAEP results are released for a variety of subjects as The Nation's Report Card, and are available for the nation, and in some cases, states and urban districts. There are no results for individual students, classrooms, or schools. NAEP reports results for different demographic groups, including gender, socioeconomic status, and race/ethnicity. Assessments are given most frequently in mathematics and reading. Other subjects, such as writing, science, U.S. history, civics, geography, economics, and the arts, are assessed periodically.

NAEP is a congressionally mandated project administered by the National Center for Education Statistics (NCES), within the Institute of Education Sciences of the U.S. Department of Education. NCES is responsible for developing test questions, administering the assessment, scoring student responses, conducting analyses of the data, and reporting the results. The National Assessment Governing Board sets policy for NAEP.
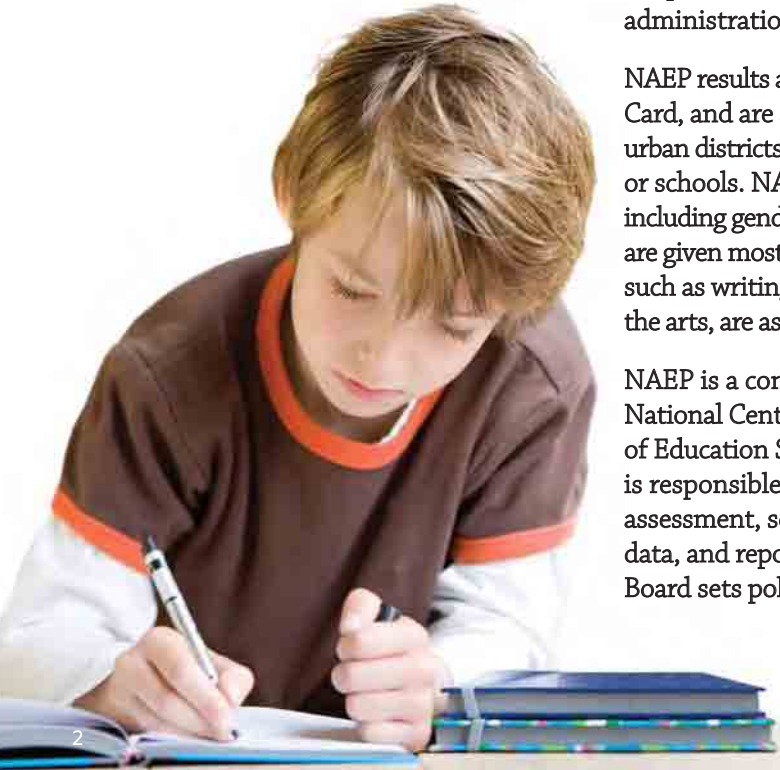
Teachers, principals, parents, policymakers, and researchers all use NAEP results to assess progress and develop ways to improve education in the United States. NAEP is a trusted resource and has been providing valid and reliable data on student performance since 1969.
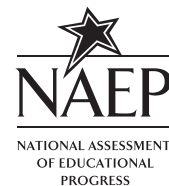
NAEP Components

There are two types of assessments, main NAEP and long-term trend NAEP.

Main NAEP is administered to fourth-, eighth-, and twelfth-graders across the country in a variety of subjects. National results are available for all assessments and subjects. Results are available for states and select urban districts in some subjects for grades 4 and 8. The Trial Urban District Assessment (TUDA) is a special project developed to determine the feasibility of reporting district-level NAEP results for large urban districts. In 2009 a trial state assessment was administered at grade 12.

Long-term trend NAEP is administered to 9-, 13-, and 17-year-olds every four years nationally. Long-term trend assessments measure student performance in mathematics and reading, and allow the performance of today's students to be compared with students since the early 1970s.

In addition to main and long-term trend assessments, there are a number of NAEP special studies that are conducted periodically. These include research and development efforts such as the High School Transcript Study and the National Indian Education Study. For more information on NAEP special studies, visit http://nces.ed.gov/nationsreportcard/studies.

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

MAIN | **LONG-TERM TREND**

National
Public & Non-public
Grades 4, 8 & 12

National
Public & Non-public
Ages: 9, 13 & 17

State
Public
Grades 4 & 8
Trial: Grade 12

Trial Urban District
Public
Grades 4 & 8
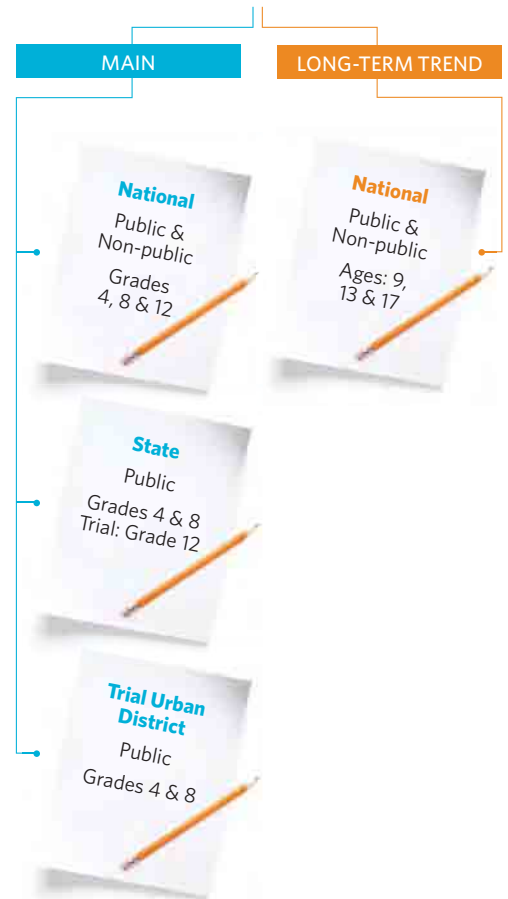
The development of a NAEP assessment takes a substantial amount of time and incorporates feedback from a wide array of experts. The process for developing NAEP assessments is comprehensive, deliberate, and inclusive.

NAEP Frameworks

Each NAEP assessment is built from a content framework that specifies what students should know and be able to do in a given grade. The National Assessment Governing Board, which sets NAEP policy, oversees the creation of the NAEP frameworks. The Governing Board, appointed by the U.S. Secretary of Education, consists of 26 members, including governors, state legislators, local and state school officials, educators, business representatives, and members of the general public.

The frameworks reflect ideas and input from subject area experts, school administrators, policymakers, teachers, parents, and others. NAEP frameworks also describe the types of questions that should be included and how they should be designed and scored. Multiple-choice and open-ended questions test a wide range of students' knowledge and skills and are presented in a variety of formats. Most assessments are administered in a paper-and-pencil based format. However, NAEP is beginning to move in the direction of computer-based assessments, starting with the writing assessment at grades 8 and 12. The frameworks are periodically updated by the Governing Board to reflect a variety of factors, such as curriculum and assessment reform. They are available online at http://www.nagb.org/.

Question Development

Once a framework is completed, NCES develops the questions that students will be asked on the assessment. More than 300 people are involved in developing and reviewing the questions for each assessment, including representatives from the framework committee, representatives from states and districts, curriculum specialists, university education faculty, state assessment directors, state content specialists, and experts in the assessment's subject area.

After the questions are written, a process of reviews and updates begins to ensure complexity, diversity of opinions and input, and attention to quality. Reviewers examine each question for:

> consistency with the framework,

> consideration for what students are being taught, and

> bias that would make a question more difficult for a particular group of students to answer.

Questions also undergo extensive pilot testing, in which actual responses from students are gathered to ensure the questions are measuring skills with fairness and appropriate levels of difficulty.

After each assessment, some NAEP questions are released to the general public. These questions are available online in the NAEP Questions Tool (http://nces.ed.gov/nationsreportcard/itmrlsx) along with data on how students performed on each question. Teachers may use these questions in their classrooms to see how their students compare to the nation, their state, or urban district (as results allow).
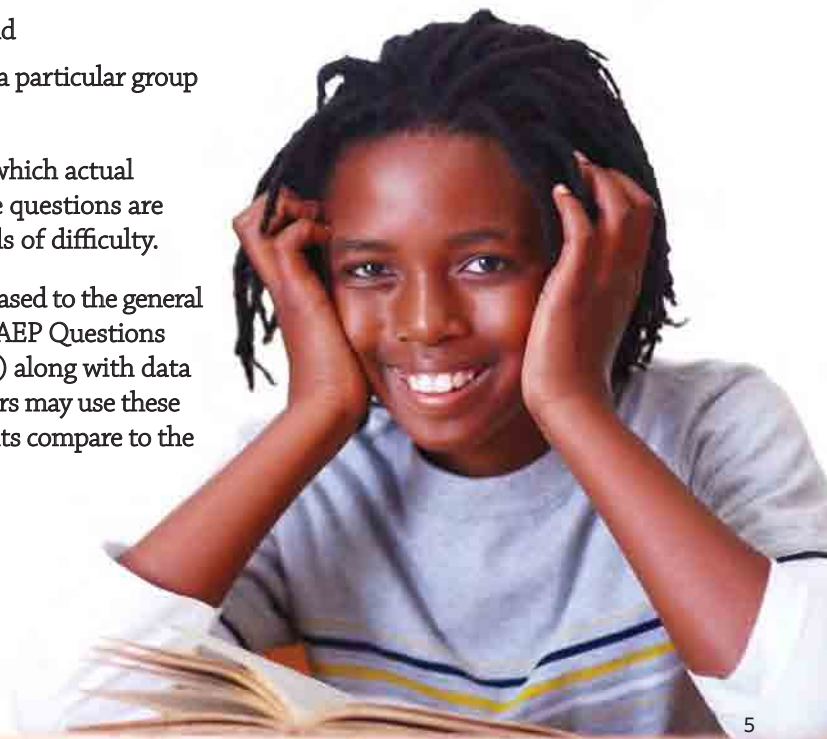
4th Grade Mathematics
Sample Question

On the grid below, plot the points that have coordinates (B, 1), (B, 3), and (D, 5).
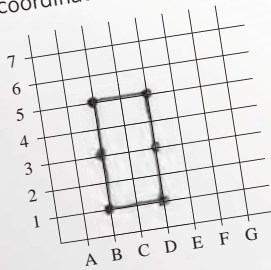
Plot 3 more points on the grid so that when you connect all 6 points you will make a rectangle.

List the coordinates for the 3 new points.

(B,5) (D,3) (D,1)

Connect the 6 points to show your rectangle.

8th Grade Reading
Sample Question

For the following question, students read a story about how a young girl's science fair project on gauging the accuracy of her city's parking meters ended up impacting the lives of many citizens throughout her state.

Choose two things Ellie Lammer did and explain what those things tell about her. Use examples from the article to support your answer.

Response rated as "Extensive"

Response rated as "Essential"

How are assessments scored?

To make sure the scoring is reliable and consistent, NCES:

> develops detailed scoring guides,
> recruits and trains qualified and experienced scorers, and
> monitors and reviews the quality and consistency of each scorer's decisions.

Computers score multiple-choice questions. Open-ended questions require written answers and a different approach to scoring. Using the scoring guides, trained scorers evaluate open-ended responses. For example, the guidelines explain how to decide whether a student's response to the question is "extended," "satisfactory," "partial," "minimal," or "incorrect." Scoring guides for released questions are available in the NAEP Questions Tool.

12th Grade Science

Sample Question

The following multiple-choice question from the twelfth-grade science assessment required conceptual understanding in the field of life science.

Which is a function of a neuron?

Ⓐ It carries oxygen to other cells.

Ⓑ It secretes digestive enzymes.

Ⓒ It removes foreign particles from the bloodstream.

⬤ It receives signals from the internal and external environments.

Questionnaires for Students, Teachers and Schools

In addition to assessing student achievement in various subjects, NAEP collects information from students, teachers, and schools in order to provide a more complete understanding of the results and overall student performance. This information is collected through the following:

> **Student questionnaires** collect information on students' demographic characteristics, classroom experiences and educational support.

> **Teacher questionnaires** gather data on teacher training and instructional practices.

> **School questionnaires** gather information on school policies and characteristics.

The results of these questionnaires help to provide contextual information for the assessments, as well as information about factors that may be related to students' learning. These results can be analyzed using the NAEP Data Explorer: http://nces.ed.gov/nationsreportcard/naepdata.
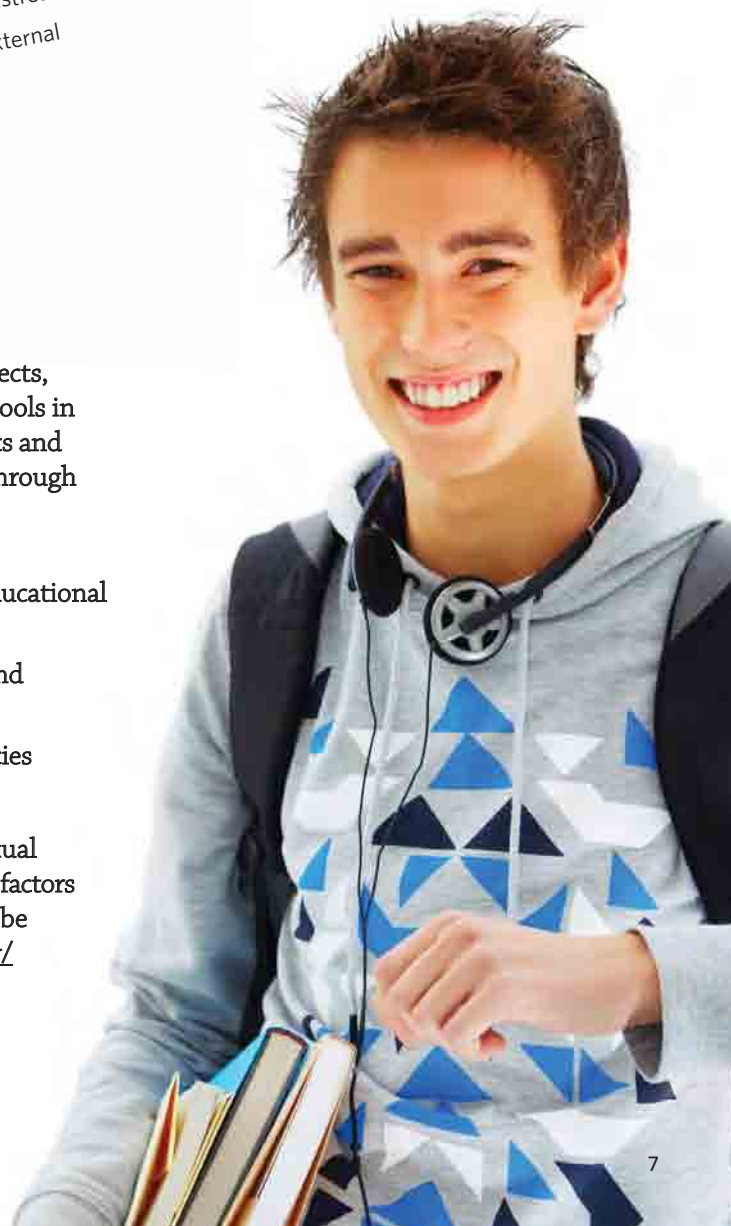
NAEP is able to provide valuable information on student performance, thanks to the voluntary participation of selected schools and students. Their participation enables NAEP to provide the most accurate and representative picture of how students are performing across the country.

Who takes NAEP?

Since NAEP does not report data for individual schools or students, it is not necessary to assess every student in the United States. NCES uses a sampling procedure to ensure that those selected to participate in NAEP will be representative of the geographical, racial, ethnic, and socioeconomic diversity of schools and students across the nation.

First, schools are selected to be representative of schools across the states and nation, on the basis of characteristics such as school location, minority enrollment, level of school achievement, and average income of the geographic area. Then, within each school, students are randomly selected to participate.

How many students participate in NAEP?

When assessing performance for the nation only, approximately 6,000 to 20,000 students per grade from across the country are assessed for each subject. The numbers of schools and students vary from year to year, depending on the number of subjects and questions to be assessed. For assessments that report state results, NAEP typically selects 3,000 students in approximately 100 schools in each state for each grade and subject.

Why is participation important?

NAEP is the only measure of how the students across the nation are performing in various subject areas. It informs us how student performance has changed over time, and allows states to compare their progress with that of other states and the nation as a whole. Without student participation, NAEP would not be as representative as it needs to be to provide fair, accurate, and useful information on student achievement.

Why are some schools selected more frequently than others?

Each year, a new sample of schools is selected. A school's participation in the past has no relation to whether it will be selected again in the future. To ensure that the student sample represents students from large and small schools in the appropriate proportions, large schools are more likely to be selected than small ones. If a school is chosen frequently, typically it is because the school's enrollment in the grade is larger than other schools in the state.

Are there accommodations for students with disabilities and for English language learners?

The goal of offering accommodations is to include as many students as possible in an assessment. This increases the representativeness of the results for students in the United States.

Test accommodations are provided for students with disabilities and English language learners who cannot fairly and accurately demonstrate their abilities under normal administration procedures. Accommodations are changes in the way a test is given, such as responding to a question in a different format or having alterations in the test setting or timing.

The most frequent accommodations used in NAEP are:

> extended time,

> small group administration,

> reading the question aloud,

> having breaks during the test administration.

However, even with accommodations, there are still some students who cannot participate in NAEP. The percentage of students included, excluded, and accommodated varies by jurisdiction. These variations should be considered when comparing student performance among the nation, states, and select urban districts.

When are the assessments given?

Main NAEP is conducted between the last week of January and the first week in March every year. Long-term trend assessments take place every four years; during the same academic year, 13-year-olds are assessed in the fall, 9-year-olds in the winter, and 17-year-olds in the spring.
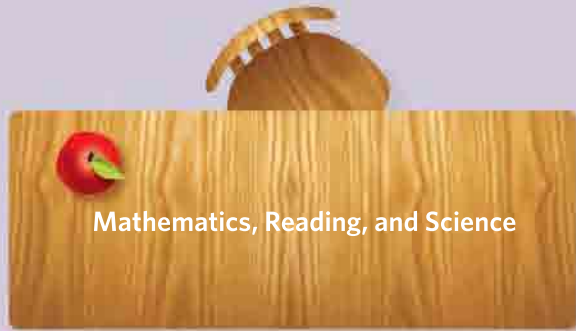
A Typical Testing Session

Every year, NAEP assessments are given in multiple subjects in the same classroom. This is a diagram of a typical classroom, with the teacher's desk at the top and 30 student desks. The diagram shows the distribution of the test booklets.

Students sitting next to each other don't take the same test. NAEP administers all the subjects together in the same classroom. This sketch of a classroom shows thirty students – the target for a NAEP session.

Mathematics, Reading, and Science

Science	Reading	Math	Science	Reading	Math
Math	Science	Reading	Math	Science	Reading
Reading	Math	Science	Reading	Math	Science
Science	Reading	Math	Science	Reading	Math
Math	Science	Reading	Math	Science	Reading

After an assessment is given and the responses are scored and analyzed, work turns to making the complex statistical results useful to the general public. NAEP results are used by policymakers, principals, teachers, and parents to inform the discussion over education.

The results of NAEP are released as The Nation's Report Card. Report cards provide results for the nation, state, and urban districts (as they are available). Results are provided as scale scores and achievement levels. The results are reported in each jurisdiction by gender, socioeconomic status, race/ethnicity, and other demographic information.

Results for the nation reflect the performance of students attending public schools, private schools, Bureau of Indian Education schools, and Department of Defense schools. Results for states and other jurisdictions reflect the performance of students in public schools only.

Scale Scores

Scale scores report what students know and can do. The results are generally reported on 0-300 or 0-500 point scales. Because NAEP scales are developed independently for each subject, scores cannot be compared across subjects.

In addition to reporting an overall score for each grade assessed, scores are reported at five percentiles to show trends in results for students performing at lower (10th and 25th percentiles), middle (50th percentile), and higher (75th and 90th percentiles) levels.

Achievement Levels

Achievement levels are standards showing what students should know and be able to do along the continuum of performance. Based on recommendations from policymakers, educators, and members of the general public, the Governing Board sets specific achievement levels for each subject and grade. NAEP results are reported as percentages of students performing at or above the *Basic* and *Proficient* levels and at the *Advanced* level.

› *Basic* denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.

› *Proficient* represents solid academic performance. Students reaching this level have demonstrated competency over challenging subject matter.

› *Advanced* represents superior performance.

Where can results be found?

NAEP results are available at http://nationsreportcard.gov. Here, users can browse or download the latest report cards, individual state and select district profiles, and access resources and online data tools that will enable them to view more extensive results, such as results by demographic groups.

Statistical Significance

Because NAEP results are based on samples, there is a margin of error surrounding each score or percentage. Differences that are larger than the margin of error are statistically significant. Significant increases or decreases from one assessment year to the next indicate that student performance has changed. Only differences that are statistically significant are discussed in NAEP reports. When interpreting NAEP results, it is important to consider whether apparent differences are statistically significant.

A number of online tools are available to explore national, state, and select urban district NAEP data for various demographic groups.

NAEP Data Explorer

The NAEP Data Explorer (NDE) is a dynamic, interactive tool used to explore assessment results for a number of various subjects, grades, and jurisdictions. It allows users to create custom statistical tables, graphics, and maps using NAEP data. Student performance in the context of gender, race/ethnicity, public or private school, teacher experience, and many other factors can be examined using data gathered from students, teachers, and schools that have participated in NAEP.

To access the NDE, visit:
http://nces.ed.gov/nationsreportcard/naepdata/

The NDE has four different databases:

› **Main NAEP** provides national results for various subjects since 1990. State and select urban district results are provided in mathematics, reading, science, and writing.

› **Long-Term Trend** provides national data on 9-, 13-, and 17-year-olds for mathematics and reading since the 1970s.

› **High School Transcript Study** provides national results for graduating seniors on NAEP assessments in mathematics and science. Results relate transcript data, such as courses taken and grade point average, to NAEP results.

› **National Indian Education Study** provides results that report on the condition of education for American Indian and Alaska Native students at grades 4 and 8 in the United States.

To access the NAEP Questions Tool (NQT), visit:
http://nces.ed.gov/nationsreportcard/itmrlsx

NAEP Questions Tool

The NAEP Questions Tool (NQT) is a database of more than 2,000 questions from past assessments, in nine subject areas, that have been released to the public and will not be used again on NAEP assessments. The NQT allows you to:

> search for questions by subject, grade, difficulty, and other characteristics,

> view student responses,

> create customized reports, and more.

You can also view scoring guides and performance data, such as the percentage of students nationwide and in your state who answered the question correctly, for most questions.

NAEP Item Maps

NAEP Item Maps display sample NAEP questions, or items, along with the scale score for students who would likely answer the question correctly. They help answer questions, such as, "What does it mean for a student to be *Basic*, *Proficient*, or *Advanced* in terms of what they know and can do?" or "What does a particular score indicate about what a student knows?"

The item maps illustrate the knowledge and skills demonstrated by students performing at different scale scores on a given assessment. They also help to provide a context for understanding NAEP results by showing concrete examples of what students at various achievement levels likely know and can do in a subject.

To access the NAEP Item Maps, visit:
http://nces.ed.gov/nationsreportcard/itemmaps/

To access the NAEP State Comparisons tool, visit:
http://nces.ed.gov/nationsreportcard/statecomparisons

NAEP State Comparisons

NAEP State Comparisons allows for sorting and comparing state results. This includes the ability to create tables and maps that compare states and jurisdictions based on the average scale scores for selected groups of public school students within a single assessment year, or compare the change in performance between two assessment years. The tool also displays whether the selected results are statistically different from one another.

State Profiles

State Profiles allows users to explore key data about a state's performance on NAEP. Users can view how a state performed over time and compare each state's overall performance to the nation and other states. Snapshot reports can also be quickly generated and downloaded. The profiles provide easy access to all NAEP data for participating states and links to the most recent state snapshots for all available subjects.
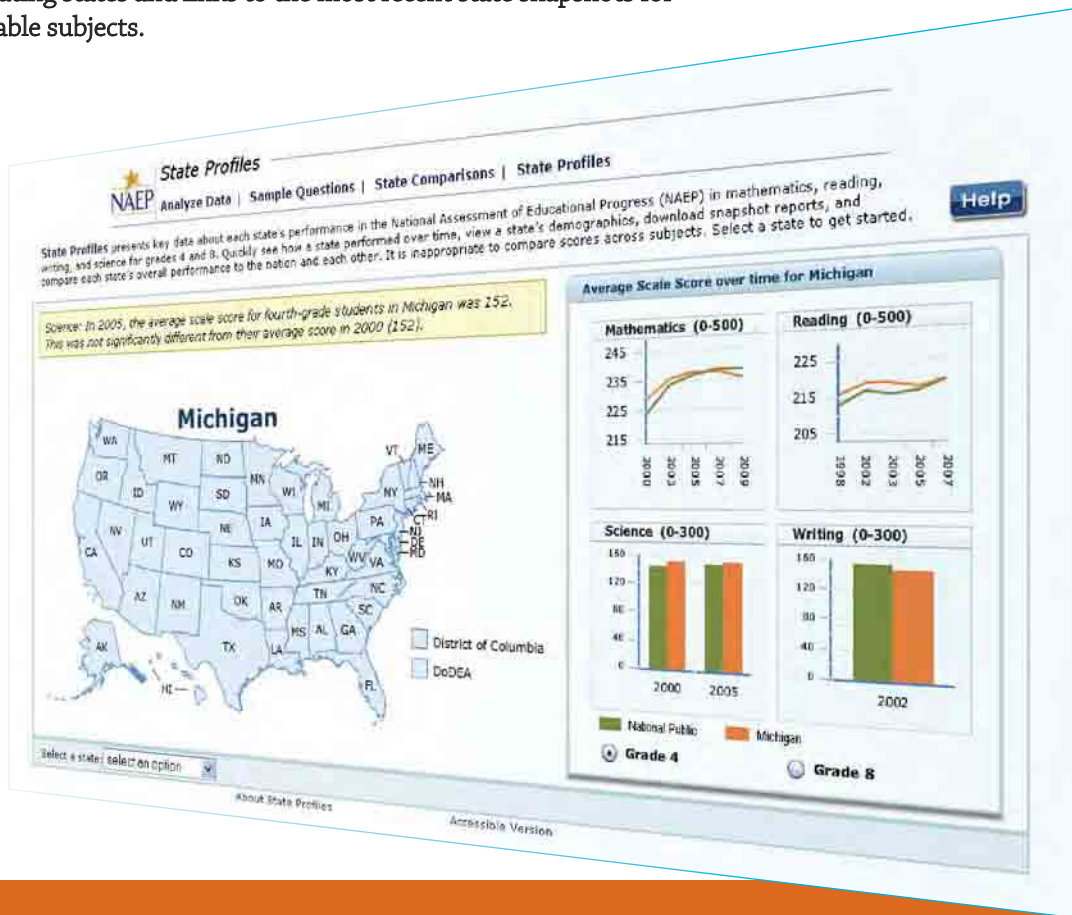
NAEP State Coordinators

NAEP is conducted in partnership with states. The NAEP program provides funding for a full-time NAEP State Coordinator (NSC) in each state. He or she serves as the liaison between NAEP, the state education agency, and the schools selected to participate.

NSCs provide many important services for the NAEP program. NSCs:

> coordinate the NAEP administration in the state,

> assist with the analysis and reporting of NAEP data, and

> promote public understanding of NAEP and its resources.

To find the name and email address of your NAEP State Coordinator, choose your state from the map at http://nces.ed.gov/nationsreportcard/states.

For more information about NAEP, visit:
http://nces.ed.gov/nationsreportcard.

You can also contact the National Center for
Education Statistics (NCES) at:

National Center for Education Statistics
Assessment Division – 8th Floor
1990 K Street NW
Washington, DC 20006
Phone: 202-502-7420
https://nces.ed.gov/nationsreportcard/contactus.asp

To order copies of The Nation's Report Card or
other NAEP publications, contact ED Pubs at:

ED Pubs
U.S. Department of Education
P.O. Box 22207
Alexandria, VA 22304
www.EDPubs.gov

www.ed.gov   ies.ed.gov

The Condition of Education 2019 At a Glance

More information is available at nces.ed.gov/programs/coe.

Preprimary, Elementary, and Secondary Education

Characteristics of Children's Families	2016	2017	Change between years
Highest level of education attained by parents of children under age 18			
Percentage whose parents' highest level of education was less than high school	10.4%	9.7%	▼
Percentage whose parents' highest level of education was a bachelor's or higher degree	39.7%	41.0%	▲
Percentage of children under age 18 living in mother-only households	26.7%	26.3%	▼
Percentage of children under age 18 in families living in poverty	19.1%	18.0%	▼

Children's Access to and Use of the Internet	2015	2017	
Percentage of children ages 3 to 18 who use the Internet from home			
Total	61%	64%	▲
Percentage of children ages 3 to 18 with no internet access at home			
Total	19%	14%	▼

Preschool and Kindergarten Enrollment	2016	2017	
Percentage of children enrolled in preprimary education			
3-year-olds	42%	40%	
4-year-olds	66%	68%	
5-year-olds	86%	86%	

Public School Enrollment	Fall 2015	Fall 2016	
Number of students enrolled in public schools	50.44 million	50.62 million	▲
Prekindergarten through 8th grade	35.39 million	35.48 million	▲
9th through 12th grade	15.05 million	15.14 million	▲

Public Charter School Enrollment	Fall 2015	Fall 2016	
Number of students enrolled in public charter schools	2.8 million	3.0 million	▲
Percentage of public school students enrolled in charter schools	5.7%	6.0%	▲
Number of public charter schools	6,860	7,010	▲
Percentage of public schools that are charter schools	7.0%	7.1%	▲

Private School Enrollment	Fall 2013	Fall 2015	
Total number of students enrolled in private schools (Prekindergarten through 12th grade)	5.4 million	5.8 million	▲
Prekindergarten through 8th grade	4.1 million	4.3 million	▲
9th through 12th grade	1.3 million	1.4 million	▲
Percentage of all students enrolled in private schools (Prekindergarten through 12th grade)	9.7%	10.2%	▲

See notes at end of table.

LEGEND: ▲ = Higher, ▼ = Lower, Blank = Not measurably different

The Condition of Education 2019

			Change between years
English Language Learners in Public Schools	Fall 2015	Fall 2016	
Percentage of public school students who are English language learners	9.5%	9.6%	▲
Children and Youth With Disabilities	2016–17	2017–18	
Number of public school students ages 3–21 receiving special education services	6.8 million	7.0 million	▲
Percentage of public school students ages 3–21 receiving special education services	13.4%	13.7%	▲
Characteristics of Traditional Public Schools and Public Charter Schools	2015–16	2016–17	
Traditional public schools			
Total number of traditional public schools	91,420	91,150	▼
Percentage of traditional public schools			
With more than 50% White enrollment	58.2%	57.4%	▼
With more than 50% Black enrollment	8.9%	8.8%	▼
With more than 50% Hispanic enrollment	16.0%	16.3%	▲
Public charter schools			
Total number of public charter schools	6,860	7,010	▲
Percentage of public charter schools			
With more than 50% White enrollment	34.4%	33.3%	▼
With more than 50% Black enrollment	23.4%	23.1%	▼
With more than 50% Hispanic enrollment	25.2%	25.8%	▲
Concentration of Public School Students Eligible for Free or Reduced-Price Lunch	2015–16	2016–17	
Percentage of students attending public low-poverty schools[1]	19.7%	21.2%	▲
Percentage of students attending public high-poverty schools[1]	24.4%	24.2%	▼
School Crime and Safety	2015	2017	
Rates of School Crime			
Percentage of students who reported criminal victimization at school	3%	2%	
Percentage of students who reported being bullied at school	21%	20%	
Characteristics of Public School Teachers	1999–2000	2015–16	
Total number of public school teachers	3.0 million	3.8 million	▲
In elementary schools	1.6 million	1.9 million	▲
In secondary schools	1.4 million	1.9 million	▲
Percentage of public school teachers			
Who are female	75%	77%	▲
Who are male	25%	23%	▼
Who held a postbaccalaureate degree	47%	57%	▲
Who held a regular teaching certificate	87%	90%	▲
	2011–12	2015–16	
Annual base salary of public school teachers[2]	$56,590	$56,140	

See notes at end of table.

LEGEND: ▲ = Higher, ▼ = Lower, Blank = Not measurably different

The Condition of Education 2019

	2011–12	2015–16	Change between years
Characteristics of Public School Principals			
Total number of public school principals	89,800	90,400	
In elementary schools	61,300	62,100	
In secondary schools	20,500	20,300	
Percentage of public school principals			
Who are female	52%	54%	▲
Who are male	48%	46%	▼
Annual base salary of public school principals[3]	$98,690	$99,670	▲

	2015	2017	
Reading Performance			
Percentage of students who scored at or above *Proficient*[4]			
4th-grade students	36%	37%	
8th-grade students	34%	36%	▲

	2013	2015	
12th-grade students	38%	37%	

	2015	2017	
Mathematics Performance			
Percentage of students who scored at or above *Proficient*[4]			
4th-grade students	40%	40%	
8th-grade students	33%	34%	

	2013	2015	
12th-grade students	26%	25%	

	2009	2015	
Science Performance			
Percentage of students who scored at or above *Proficient*[4]			
4th-grade students	34%	38%	▲
12th-grade students	21%	22%	

	2011	2015	
8th-grade students	32%	34%	

	2015–16	2016–17	
Public High School Graduation Rates			
Adjusted Cohort Graduation Rate (ACGR)[5]	84%	85%	▲

	2016	2017	
Status Dropout Rates			
Percentage of 16- to 24-year-olds not enrolled in school who have not completed high school	5.8%	5.4%	▼
Male	6.8%	6.4%	▼
Female	4.7%	4.4%	▼
White	4.5%	4.3%	▼
Black	7.0%	6.5%	▼
Hispanic	9.1%	8.2%	▼
Asian	2.0%	2.1%	
Pacific Islander	6.9%	3.9%	▼
American Indian/Alaska Native	11.0%	10.1%	
Two or more races	4.8%	4.5%	

See notes at end of table.

LEGEND: ▲ = Higher, ▼ = Lower, Blank = Not measurably different

The Condition of Education 2019

Public School Revenue Sources[3]	2014–15	2015–16	Change between years
Total revenues	$679.0 billion	$706.4 billion	▲
Federal sources	$57.7 billion	$58.3 billion	▲
State sources	$316.1 billion	$331.7 billion	▲
Local sources	$305.2 billion	$316.4 billion	▲
Public School Expenditures[3]	**2014–15**	**2015–16**	
Total expenditures	$683 billion	$706 billion	▲
Current expenditures per student	$11,998	$12,330	▲

Postsecondary Education

Immediate College Enrollment Rate	2016	2017	Change between years
Percentage of recent high school graduates enrolled in college	70%	67%	
2-year institutions	24%	23%	
4-year institutions	46%	44%	
College Enrollment Rates	**2016**	**2017**	
College participation rates for 18- to 24-year-olds			
Total, all students	41%	40%	
Female	44%	44%	
Male	39%	37%	
White	42%	41%	
Black	36%	36%	
Hispanic	39%	36%	
Asian	58%	65%	▲
Pacific Islander	21%	33%	
American Indian/Alaska Native	19%	20%	
Two or more races	42%	41%	
Undergraduate Enrollment	**Fall 2016**	**Fall 2017**	
Total enrollment	16.87 million	16.76 million	▼
Full-time enrollment	10.43 million	10.37 million	▼
Part-time enrollment	6.44 million	6.39 million	▼
Percentage enrolled in any distance education course	30.8%	32.9%	▲
Percentage enrolled exclusively in distance education	12.8%	13.3%	▲
Postbaccalaureate Enrollment	**Fall 2016**	**Fall 2017**	
Total enrollment	2.97 million	3.01 million	▲
Full-time enrollment	1.70 million	1.71 million	▲
Part-time enrollment	1.28 million	1.30 million	▲
Percentage enrolled in any distance education course	32%	34%	▲
Percentage enrolled exclusively in distance education	15%	16%	▲

See notes at end of table.

LEGEND: ▲ = Higher, ▼ = Lower, Blank = Not measurably different

Characteristics of Postsecondary Students	2016–17	2017–18	Change between years
Total enrollment	19.84 million	19.77 million	▼
Undergraduate enrollment	16.87 million	16.76 million	▼
White	9.08 million	8.88 million	▼
Black	2.23 million	2.18 million	▼
Hispanic	3.17 million	3.27 million	▲
Asian	1.05 million	1.07 million	▲
Pacific Islander	47,100	46,100	▼
American Indian/Alaska Native	128,600	124,000	▼
Two or more races	595,700	623,400	▲
Nonresident alien	570,300	575,000	▲
Postbaccalaureate enrollment	2.97 million	3.01 million	▲
White	1.63 million	1.63 million	▲
Black	362,900	365,400	▲
Hispanic	259,600	275,000	▲
Asian	200,200	208,900	▲
Pacific Islander	6,100	5,900	▼
American Indian/Alaska Native	13,700	13,600	▼
Two or more races	70,700	76,800	▲
Nonresident alien	427,800	425,700	▼

Characteristics of Degree-Granting Postsecondary Institutions	2016–17	2017–18	
Total number of degree-granting institutions with first-year undergraduates	3,895	3,883	▼
Number of 4-year institutions with first-year undergraduates	2,395	2,407	▲
Number of 2-year institutions with first-year undergraduates	1,500	1,476	▼

Characteristics of Postsecondary Faculty	Fall 2016	Fall 2017	
Number of full-time instructional faculty[6]	814,000	821,000	▲
Number of part-time instructional faculty	732,000	722,000	▼

Undergraduate Degree Fields	2015–16	2016–17	
Number of bachelor's degrees conferred			
Business	371,700	381,400	▲
Health professions and related programs	228,900	238,000	▲
Social sciences and history	161,200	159,100	▼

Graduate Degree Fields	2015–16	2016–17	
Number of master's degrees conferred			
Business	186,800	187,400	▲
Education	145,800	145,700	▼
Health professions and related programs	110,400	119,300	▲

See notes at end of table.

LEGEND: ▲ = Higher, ▼ = Lower, Blank = Not measurably different

Undergraduate Retention and Graduation Rates	2015–16	2016–17	Change between years
4-year institutions			
Retention rate of first-time undergraduates	80.8%	81.0%	▲
Graduation rate (within 6 years of starting program) of first-time, full-time undergraduates	59.7%	60.4%	▲
2-year institutions			
Retention rate of first-time undergraduates	62.3%	62.5%	▲
Graduation rate (within 150% of normal time for degree completion) of first-time, full-time undergraduates	30.3%	31.6%	▲
Postsecondary Certificates and Degrees Conferred	**2015–16**	**2016–17**	
Number of degrees/certificates conferred by postsecondary institutions			
Certificates below associate's degrees	939,000	945,000	▲
Associate's degrees	1,008,000	1,006,000	▼
Bachelor's degrees	1,921,000	1,956,000	▲
Master's degrees	786,000	805,000	▲
Doctor's degrees	178,000	181,000	▲
Price of Attending an Undergraduate Institution[3]	**2015–16**	**2016–17**	
Average net price at 4-year institutions for first-time, full-time undergraduate students			
Public, in-state or in-district[7]	$13,660	$13,760	▲
Private nonprofit	$26,780	$26,840	▲
Private for-profit	$22,660	$22,000	▼
Loans for Undergraduate Students	**2015–16**	**2016–17**	
Percentage of undergraduates with student loans	45.6%	46.1%	▲
Average student loan amount[3]	$7,280	$7,240	▼
Sources of Financial Aid	**2015–16**	**2016–17**	
Percentage of students receiving any financial aid at 4-year institutions	84.9%	85.0%	▲
Percentage of students receiving any financial aid at 2-year institutions	77.5%	77.8%	▲
Postsecondary Institution Revenues[3]	**2015–16**	**2016–17**	
Revenue from tuition and fees per full-time-equivalent (FTE) student			
Public institutions	$7,547	$7,666	▲
Private nonprofit institutions	$21,872	$21,881	▲
Private for-profit institutions	$16,315	$16,474	▲
Postsecondary Institution Expenses[3]	**2015–16**	**2016–17**	
Instruction expenses per full-time-equivalent (FTE) student			
Public institutions	$10,670	$10,832	▲
Private nonprofit institutions	$18,270	$18,384	▲
Private for-profit institutions	$4,474	$4,483	▲

See notes at end of table.

LEGEND: ▲ = Higher, ▼ = Lower, Blank = Not measurably different

Population Characteristics and Economic Outcomes

Educational Attainment of Young Adults	2017	2018	Change between years
Percentage of 25- to 29-year-olds with selected levels of educational attainment			
High school completion or higher	92%	93%	
Associate's or higher degree	46%	47%	
Bachelor's or higher degree	36%	37%	
Master's or higher degree	9%	9%	

Young Adults Neither Enrolled in School nor Working	2016	2017	
Percentage of 18- to 24-year-olds who were neither enrolled in school nor working			
Total 18- to 24-year-olds	14%	14%	
18- and 19-year-olds	10%	11%	
20- to 24-year-olds	16%	15%	▼
White	12%	11%	▼
Black	21%	22%	
Hispanic	17%	16%	
Asian	8%	7%	
Pacific Islander	16%	20%	
American Indian/Alaska Native	32%	29%	
Two or more races	14%	14%	

Annual Earnings of Young Adults	2016	2017	
Median annual earnings for 25- to 34-year-olds[3]			
Total	$40,900	$41,900	▲
With less than high school completion	$25,900	$26,000	
Who completed high school as highest level	$32,500	$32,000	
Who completed some college but did not attain a degree	$35,600	$35,000	
Who attained an associate's degree	$38,800	$38,900	
Who attained a bachelor's or higher degree	$55,900	$55,000	
Who attained a bachelor's degree	$51,100	$51,800	
Who attained a master's or higher degree	$65,400	$65,000	

Employment and Unemployment Rates by Educational Attainment	2017	2018	
Employment rates of 25- to 34-year-olds			
Total	78%	79%	
With less than high school completion	57%	59%	
Who completed high school as highest level	72%	72%	
Who attained a bachelor's or higher degree	86%	86%	
Unemployment rates of 25- to 34-year-olds			
Total	5%	4%	▼
With less than high school completion	13%	9%	▼
Who completed high school as highest level	7%	6%	
Who attained a bachelor's or higher degree	3%	2%	

See notes at end of table.

LEGEND: ▲ = Higher, ▼ = Lower, Blank = Not measurably different

The Condition of Education 2019

International Comparisons

International Comparisons: Reading Literacy at Grade 4 (2016)	U.S. average score	International average score	Difference between the U.S. average and the international average
Progress in International Reading Literacy Study (PIRLS)			
Average reading literacy scores of 4th-grade students	549	500	▲
Average online informational reading score of 4th-grade students	557	500	▲

International Comparisons: U.S. 4th-, 8th-, and 12th-Graders' Mathematics and Science Achievement (2015)	U.S. average score	TIMSS scale center-point	Difference between the U.S. average and the TIMSS scale center-point
Trends in International Mathematics and Science Study (TIMSS)			
Mathematics scores of 4th-grade students	539	500	▲
Mathematics scores of 8th-grade students	518	500	▲
Science scores of 4th-grade students	546	500	▲
Science scores of 8th-grade students	530	500	▲
TIMSS Advanced			
Advanced mathematics scores of 12th-grade students	485	500	▼
Physics scores of 12th-grade students	437	500	▼

International Comparisons: Science, Reading, and Mathematics Literacy of 15-Year-Old Students (2015)	U.S. average score	OECD average score	Difference between the U.S. average and the OECD average
Program for International Student Assessment (PISA)			
Science literacy scores of 15-year-old students	496	493	
Reading literacy scores of 15-year-old students	497	493	
Mathematics literacy scores of 15-year-old students	470	490	▼

Enrollment Rates by Country	2015	2016	Change between years
Percentage of 3- and 4-year-olds enrolled in school at any level			
United States	54.4%	52.7%	▼
Organization for Economic Cooperation and Development (OECD) countries	81.0%	81.9%	▲
Percentage of 5- to 14-year-olds enrolled in school at any level			
United States	98.0%	99.2%	▲
OECD countries	98.2%	98.1%	▼

See notes at end of table.

LEGEND: ▲ = Higher, ▼ = Lower, Blank = Not measurably different

			Change between years
Percentage of 15- to 19-year-olds enrolled in school at any level			
United States	81.5%	82.5%	▲
OECD countries	84.6%	84.9%	▲
Percentage of 20- to 29-year-olds enrolled in school at any level			
United States	24.9%	24.5%	▼
OECD countries	28.7%	28.6%	▼

International Educational Attainment	2016	2017	
Percentage of the population 25 to 34 years old who completed high school			
United States	91.5%	92.1%	▲
Organization for Economic Cooperation and Development (OECD) countries	84.3%	84.8%	▲
Percentage of the population 25 to 34 years old who attained a postsecondary degree			
United States	47.5%	47.8%	
OECD countries	43.4%	44.5%	▲

Education Expenditures by Country (2015)[8]	U.S.	OECD	Difference between the U.S. and OECD
Expenditures per full-time-equivalent (FTE) student			
Elementary and secondary education	$12,800	$9,500	▲
Postsecondary education	$31,000	$16,100	▲

[1] Low-poverty schools are defined as public schools where 25 percent or less of the students are eligible for free or reduced-price lunch (FRPL). A high-poverty school is defined as a public school where more than 75 percent of the students are eligible for FRPL.
[2] Data are reported in constant 2016–17 dollars, based on the Consumer Price Index (CPI).
[3] Data are reported in constant 2017–18 dollars, based on the Consumer Price Index (CPI).
[4] *Proficient* demonstrates solid academic performance and competency over challenging subject matter.
[5] The Adjusted Cohort Graduation Rate (ACGR) is the number of students who graduate in 4 years with a regular high school diploma divided by the number of students who form the adjusted cohort for the graduating class. From the beginning of 9th grade (or the earliest high school grade), students who enter that grade for the first time form a cohort that is "adjusted" by adding any students who subsequently transfer into the cohort and subtracting any students who subsequently transfer out, emigrate to another country, or die.
[6] Data are for full-time instructional faculty on 9-month contracts at degree-granting postsecondary institutions.
[7] The average net price at public 4-year institutions uses the lower of in-district or in-state average net price.
[8] Data are reported in constant 2017 dollars based on the OECD's National Consumer Price Index.
NOTE: All calculations within the At a Glance are based on unrounded numbers. Race categories exclude persons of Hispanic ethnicity.
SOURCE: *The Condition of Education 2019.*

LEGEND: ▲ = Higher, ▼ = Lower, Blank = Not measurably different

HOW DID U.S. STUDENTS PERFORM ON THE MOST RECENT ASSESSMENTS?

SHARE

Select a **JURISDICTION** and a **RESULT** to see how students performed on the latest NAEP assessments.
Click on the ▼ to see the most recent reports in each subject.

SELECT JURISDICTION:
Public and nonpublic schools ▼

SELECT RESULT:
At or above *NAEP Proficient*

PERCENTAGE OF STUDENTS AT OR ABOVE *NAEP Proficient*

	Grade 4	Grade 8	Grade 12
ARTS: MUSIC	—	—	—
ARTS: VISUAL ARTS	—	—	—
CIVICS	27	23	24
ECONOMICS	—	—	42
GEOGRAPHY	21	27	20
MATHEMATICS	41	34	25
READING	35	34	37
SCIENCE	38	34	22
TECHNOLOGY & ENGINEERING LITERACY	—	46	—
U.S. HISTORY	20	18	12
WRITING	28	27	27

KEY: ▬ DATA NOT AVAILABLE

NOTE: DoDEA – Department of Defense Education Activity (overseas and domestic schools). Year shown with achievement-level results is the most recent assessment year in that grade and subject combination. Not all assessments are administered at all grades. For more information on each subject, click on the report of interest in the menu at the top of this page under the report tab. Base year indicates the first assessment year in that grade and subject combination.

The Nation's Report Card
is the largest ongoing assessment of what U.S. students know and can do.

NEW! **Puerto Rico mathematics results now available**

Results of the 2019 mathematics and reading assessments

Have students made progress since 2017?

- Higher average score in grade 4 mathematics
- Lower average score in grade 8 mathematics
- Lower average reading scores at both grades

| Mathematics Highlights | Reading Highlights |

Download a **summary of results**.

NAEP Validity Studies Panel Responses to the Reanalysis of TUDA Mathematics Scores

This report explores the consequences of differences in mathematics content emphasis between NAEP and selected state assessments for estimating score means for TUDAs.

National Indian Education Study: A Closer Look

The latest NIES report takes a closer look at how American Indian/Alaska Native students see themselves in terms of their Native languages and culture. This companion to the **2015 NIES report** also discusses student engagement at school and expectations for the future.

Upcoming Reports

- **2018 Civics, Geography, and U.S. History**
- **2019 Mathematics and Reading Report Cards at Grade 12**

Focus on NAEP

National Assessment of Educational Progress

Sampling

Focus on NAEP is a publication series developed to highlight findings and summarize information about the ongoing development and implementation of the National Assessment of Educational Progress (NAEP). This issue details how samples of schools and students are selected for the NAEP assessments.

National

National samples include both public and non-public schools. For national-only NAEP assessments (e.g., Long-Term Trend), sampling begins with primary sampling units (PSUs) from the U.S. Census Bureau, which are a merging of the nation's 3,000 counties. PSUs are categorized based on geographical (e.g., census region and locale) and demographic factors (e.g., income level).

Ten percent of these PSUs are selected using probability proportional to sample size, meaning that larger PSUs have a greater likelihood of selection, resulting in a sample of 95 to 100 PSUs. Sampling in certain NAEP assessments (i.e., reading and mathematics) has expanded to allow for state-level analysis. In these assessments public school data for the national sample is composed of the sum of the state samples.

State

In the 1990s, the NAEP program expanded the assessment to the 50 states, the District of Columbia, and the Department of Defense Education Activity (DoDEA). State samples include only public schools. In a NAEP state assessment year the entire state is treated as a sampling unit.
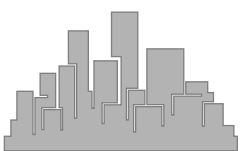
TUDA Districts

In 2002, the first NAEP Trial Urban District Assessment (TUDA) was conducted in six large urban districts. Currently there are 21 TUDA districts. Larger TUDA districts have somewhat larger sample sizes than others. All data collection for TUDA districts contributes to state results.

NAEP Representative Samples

NAEP is designed to report results at the national and state level, as well as for selected urban districts, without requiring every student in every school to take the assessment. This is accomplished by creating a sampling frame using the Common Core of Data (CCD) and the Private School Survey (PSS). NAEP is then administered to a sample of students from this frame who represent the student population of the nation as a whole and of individual states and districts participating in the Trial Urban District Assessment (TUDA).

School Selection

1 Identify schools

Schools are identified based on the U.S. Department of Education's public school system database. Because national-level data on schools can be up to three years old, school data is verified with state departments of education. A national sample of nonpublic (private) schools is also selected for grades 4, 8, and 12.

Classify schools 2

Schools are classified into groups by type of location (city, suburb, town, or rural) and then by racial/ethnic composition. These groups are called strata.

3 Arrange schools

Schools are sorted by a student achievement measure (e.g., school-level results on state achievement tests) to ensure NAEP represents all levels of school performance.

4 List schools

The school groupings are then placed into a comprehensive ordered list. The specific schools that comprise the sample are ordered based on location type, race/ethnicity, and student achievement.

Select schools 5

NCES draws a separate sample of schools from each stratum with probability proportional to school size. Small schools, high minority schools, and private schools are sampled to ensure that they are adequately represented.

If the school participation rate in the sample is below 85 percent, the results cannot be reported.

6 Confirm schools

A list of selected schools is sent to each state department of education to confirm eligibility. School closure or no eligible students would prevent a school from being selected.

Student Selection

Once a school has been selected for either a state or national assessment, students within the school are selected for each subject area based on grade (4, 8, or 12). From this list, a sample of students is randomly selected by NCES to participate in the assessment. Every student in a sampled school is eligible to be selected.

After the sample is drawn, students are assigned to a single subject area in which to answer questions. In general, 30 students are selected per grade per subject from each school. Typically, 95 percent of fourth- and eighth-grade students and 85 percent of twelfth-grade students selected agree to participate.

Students with disabilities and English language learners are eligible to receive accommodations offered by the NAEP program.

Grade 4   Grade 8   Grade 12
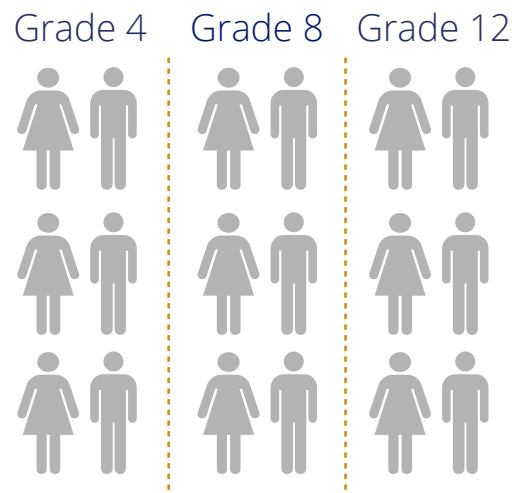
Subject Selection

This diagram shows the typical distribution of NAEP test booklets in a classroom of 30 students during an assessment of mathematics, reading, and science. Students sitting next to each other do not take the same subject. For each subject, there are several different versions of booklets.

Reading	Science	Math	Reading	Science	Math
Math	Reading	Science	Math	Reading	Science
Science	Math	Reading	Science	Math	Reading
Reading	Science	Math	Reading	Science	Math
Math	Reading	Science	Math	Reading	Science

Item Selection

Individual students only complete a subset of NAEP items. A technique called Balanced Incomplete Block (BIB) spiraling is used in which groups of items are systematically arranged with other groups of items within test booklets to assure that the entire content domain for the subject is covered. And all items are completed by a representative sample of students, though individual students only complete a fraction of the items.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013 Mathematics and Reading Assessments.

NAEP Technical Documentation Website

NAEP Technical Documentation NAEP 2013 Sample Design

The sample design for NAEP 2013 included samples for various operational, special study, and pilot test assessments. Representative samples were drawn for the following operational assessments:

2013 State Assessment Sample Design

2013 National Assessment Sample Design

- national assessments in mathematics and reading in public and private schools at grades 4, 8, and 12;
- state-by-state and Trial Urban District Assessments (TUDA) assessments in mathematics and reading in public schools at grades 4 and 8; and
- state-by-state assessments in mathematics and reading in public schools at grade 12 in 13 states.

Representative samples were drawn for the following special studies and pilot test assessments:

- pilot test of the computer-based assessment of Technical and Engineering Literacy (TEL) in public schools at grade 8;
- a special mathematics assessment in public and private schools in Puerto Rico at grades 4 and 8;
- Accessible Booklet Study in reading in public and private schools at grades 4 and 8;
- study to examine the link between Lexile and NAEP reading in public and private schools at grade 8;
- study to obtain NAEP grade 12 mathematics scores for students in the National High School Longitudinal Study (HSLS) in public schools;
- study to examine the relationship between NAEP grade 8 and grade 12 mathematics scales, conducted in public schools at grades 9, 10, and 11 in two states that conducted PISA assessments in 2012; and
- pilot tests in reading and mathematics in public and private schools at grades 4 and 8.

The samples for the operational assessments were organized into four distinct components and sampled separately. The samples for the special studies and pilot tests were integrated into these various components:

- mathematics and reading assessments in public schools at grades 4 and 8;
- mathematics and reading assessments in public schools at grade 12;
- mathematics and reading assessments in private schools at grades 4, 8, and 12; and
- computer-based TEL pilot assessment in public schools at grade 8.

The national assessments were designed to achieve nationally representative samples of public and private school students in the fourth, eighth, and twelfth grades. Their target populations included all students in public, private, Bureau of Indian Education (BIE), and Department of

Defense Education Activity (DoDEA) schools, who were enrolled in grades 4, 8, and 12 at the time of assessment.

For the fourth- and eighth-grade mathematics and reading assessments in public schools, the NAEP state student samples and assessments constituted the NAEP national student samples and assessments. Nationally representative samples were drawn for the remaining populations of private school students, DoDEA students, and BIE students in the fourth and eighth grades.

The TUDA samples formed part of the corresponding state public school samples, and the state samples formed the public school grades 4 and 8 part of the national sample.

At grade 12, the national samples for mathematics and reading consisted of 13 state samples of public schools and additional samples of public, private, BIE, and DoDEA schools to represent the balance of the nation.

All samples except the TEL pilot sample were based on a two-stage sample design:

- selection of schools within strata; and
- selection of students within schools.

The computer-based TEL pilot sample was based on a three-stage sample design:

- selection of primary sampling units (PSUs);
- selection of schools within strata; and
- selection of students within schools.

In the three-stage design for the TEL pilot sample, schools were stratified and selected within the sampled PSUs. The sample of schools was selected with probability proportional to a measure of size based on the estimated grade 8 student enrollment.

The state assessments were designed to achieve representative samples of students in the respective grade. At grades 4 and 8, the target populations included all students in each participating jurisdiction, which included states, District of Columbia, DoDEA, and school districts chosen for the TUDA assessments. At grade 12, the target population consisted of all students in each of the 13 participating states. Each sample was designed to produce aggregate estimates with reliable precision for all the participating jurisdictions, as well as estimates for various student subpopulations of interest.

In the PISA linking study, samples of students in grades 9 through 11 were selected from the schools selected for the grade 12 public school samples in Florida and Massachusetts.

The figure below illustrates the various sample types and subjects.

Components of the NAEP samples, by assessment subject, grade, and school type: 2013

Grade	Assessment		TEL Pilot
	Reading	Mathematics	
4	Public/BIE/DoDEA		
	Private		
8	Public/BIE/DoDEA		Public
	Private		
12	Public/BIE/DoDEA		
	Private		

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013 Assessments.

http://nces.ed.gov/nationsreportcard/tdw/sample_design/2013/naep_2013_sample_design.aspx
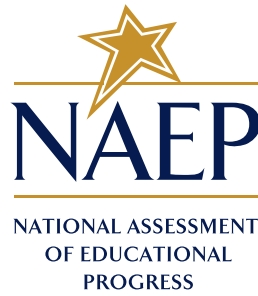
FAQs: NAEP State Assessment Sample Design

How are schools and students selected for NAEP?

NAEP provides results on student achievement, instructional experiences, and school environment factors for the nation, states, and participating urban districts. Since NAEP is not designed to report results for individual students or schools, it is not necessary for every student in every school to take the assessment. Instead, an accurate picture of student performance is obtained by administering NAEP to a sample of students who represent the student population of the nation as a whole and of individual states and districts participating in TUDA.

To ensure that a representative sample of students is assessed, NAEP is given in a sample of schools whose students reflect the varying demographics of a specific jurisdiction, be it the nation, a state, or a district. Within each selected school and grade to be assessed, students are chosen at random to participate in NAEP. Every student has the same chance of being chosen—regardless of race/ethnicity, socioeconomic status, disability, status as an English language learner, or any other factors.

It is important that all selected schools and students participate in NAEP. Participation is vital for valid information about the performance of students across the country to be collected and shared. Elected officials, policymakers, and educators all use NAEP results to develop ways to improve education.

The following steps are used to select a sample of public schools and students in a year when NAEP reports state-level results. Private schools are not included in a state-level sample, which focuses solely on public schools.

1. Identify all potential schools in each state.

An initial list of all public schools in the nation is compiled using the U.S. Department of Education's most current public education system database. The national list is then divided into separate lists of schools within each state to begin the sampling proces

Identify Potential Schools

2. Classify schools into groups.

Using the list, schools within each state are classified into groups, first by type of location and then by the racial/ethnic composition of the schools within those locations. This step ensures that the sampling process takes into account the distribution of schools and students across rural, suburban, and urban areas, and the diversity of the student population in each state.

Classify Schools Into Groups

Group A   Group B
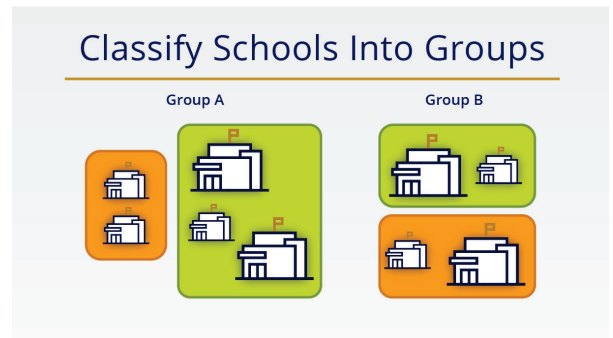
3. Within each group, order schools by student achievement.

In each sampled school, a list is compiled of all students within the grade to be assessed. From this list, a sample of students is randomly selected by NCES for participation in the assessment. Every student in a sampled school has an equal probability of being selected. After the sample is drawn, students are assigned a single subject area in which to answer questions. NAEP staff members work with the school to verify the accuracy of student demographic information

Order Schools By Achievement

Group A   Achievement

0 100

0 100

4. Develop an ordered list for sampling.

The groupings of schools determined in steps 2 and 3 are then placed into a comprehensive ordered list for sampling, i.e. by type of location, race/ethnicity, and student achievement. The probability of a school being selected for the NAEP sample is calculated based on the size of its enrollment in relation to the size of the state's student population at the selected grade level and the number of schools needed for the assessment. Schools with large enrollments are more likely to be selected because their students represent a large proportion of the state's student population.represent a large proportion of the state's student population.

Develop Ordered Sample

5. Select the school sample.

The sample of schools is drawn for NAEP participation with a systematic sampling procedure. The procedure ensures that each school has the correct selection probability, as calculated in Step 4. By proceeding systematically throughout the entire list, schools of different sizes and varying demographics are selected and a representative sample of students in the state will be chosen for the assessment.

Select School Sample

6. Confirm school eligibility.

The list of schools selected to participate is sent to each state department of education to verify that the schools are eligible for participation. Some factors that would make a school ineligible include school closure or if the school does not have students in the grade level being assessed.

Confirm School Eligibility

7. Within sampled schools, select students to participate in NAEP.

The National Assessment of Educational Progress (NAEP) is the largest nationally representative and continuing assessment of what America's students know and can do in various subject areas. Paper-and-pencil assessments are conducted periodically in mathematics, reading, science, writing, the arts, civics, economics, geography, U.S. history, and in Technology and Engineering Literacy (TEL). Beginning in 2017, NAEP will begin administering digitally based assessments (DBA) for mathematics, reading, and writing, with additional subjects added in 2018 and 2019.

Select Students to Participate
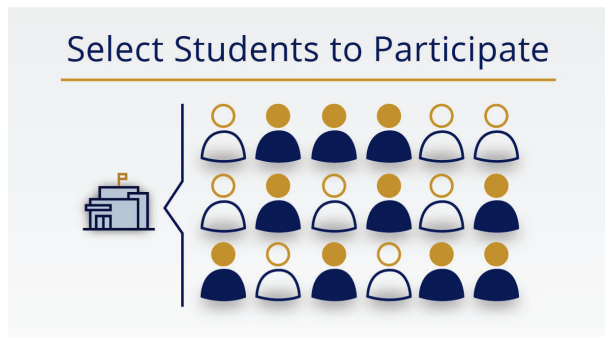
NAEP assessments are conducted across the nation, states, and in some cases, urban districts. National results are reported for all assessment subjects. Mathematics, reading, science, and writing are also assessed at the state level and for participating urban districts. For national assessments, students in both public and private schools are assessed. At the state level, only public school assessments are reported. Reports are never reported for individual students.

Why are some schools selected frequently?

The probability (or chance) of being selected for the NAEP state sample depends on the size of the student enrollment at each school. Generally, if a school is chosen frequently it is because the school's enrollment in the grade constitutes a relatively large proportion of the state's student population. Therefore, large schools are more likely to be selected than smaller ones. The sampling process is repeated each year that NAEP is administered, each time using the most recent data to account for changes in schools and shifts in student demographics across states and the nation.

For state assessments, NAEP selects for the sample around 100 public schools for each subject at each grade—each school represents about 1% of that state's public school students in the grade being assessed. If a school is chosen repeatedly, typically that is because their enrollment in the grade represents more than 1% of the state's enrollment in the grade. Other schools, with 0.5%-1% of the enrollment, are not always selected, but are selected

very frequently.

Are any schools or students targeted for oversampling?

For NAEP assessments based on state samples (mathematics, reading, science, and writing at grades 4 and 8), the samples within each state are designed (with one exception mentioned below) to be proportionately representative of all the different groups of students in the state. No students are targeted for oversampling based on their student characteristics. However, in a state that contains one or more districts that are participating in the Trial Urban District Assessment (TUDA), students from the districts involved are sampled at a greater rate than those in the remainder of the state. NAEP's weighting procedures ensure that the final results for the state are based on the correct proportional contribution from these TUDA districts despite this difference in sampling rates within the state. National samples for other subjects, for grade 12, and for private schools may require oversampling.

Why do the numbers of schools and students selected to participate vary from year to year?

The number of schools and students selected to participate in NAEP varies each year, depending on the number of subjects being assessed and the level of results to be reported (i.e., national, state, or district-level).

National Results: For NAEP results to be representative of the nation, far fewer schools and students need to participate than when results for states are being reported. Within each state, the number of schools selected to participate will vary depending on the size of its student population. For example, California has a much larger share of the student population, so it is much more likely to have schools in the sample than Wyoming.

State Results: In years when NAEP also reports state-level results, larger numbers of schools are needed to ensure that the results are representative of each state's student population. Approximately 3,000 public school students per subject and grade are assessed in each state.

District Results: The number of schools and students needed to represent the districts participating in TUDA varies, depending on the size of the district. Generally, in each district, approximately 1,500 public school students per subject and grade are assessed. These schools and students are added to and contribute to the state's sample. While TUDA districts have a large sample, their data are combined with the state's data proportionally so as not to over-represent the district in the overall state results.

On the Nation's Report Card website, you can see rounded numbers of the schools and students that participated in recent assessments.

How are private schools selected to participate in NAEP?

The process for private school selection is similar to the public school selection process, but depends on the U.S. Department of Education's private education system databases to create the initial list of all known private schools. Private schools are sampled to be representative of private schools nationwide. The results for private schools are not included in state-level results which are solely focused on public schools.

Can states choose which schools are sampled for NAEP?

No. NCES selects a sample of schools using the NAEP sampling process to ensure the validity of the results. States verify if the selected schools are eligible to participate. This process ensures that NAEP assesses the most representative sample of students as possible.

How do school response rates affect NAEP reporting?

The National Center for Education Statistics and the National Assessment Governing Board have established participation rate standards that states and jurisdictions are required to meet in order to have their results published. Beginning in 2003, if a state's school response rate is below 85%, the results will not be published by NAEP, regardless of the response rate after substitution. To read details about response rates for a particular assessment, see the section on weighting procedures in the technical documentation. From this page, look to the right, and choose the year you are interested in. On the weighting procedures page for the year selected, find the link "Quality Control…" on the right, and from the page that appears, click "Nonresponse Bias Analyses."

Are students with disabilities and English language learners included in the NAEP sample?

Yes. Students are randomly selected from a list of all students enrolled in the grade that is to be assessed in each school. NAEP's policy has always been to include as many sampled students as possible in the assessment so that results are representative of all students. NAEP offers a comprehensive array of research-based accommodations for students with disabilities, students with 504 plans, and English language learners to participate in the assessment.

How many students participate in each NAEP assessment?

For the most recent assessments, find out how many students participate in the FAQs on The Nation's Report card website. For assessments as early as 2000, to find numbers of schools and students as well as participation rates, go to NAEP technical documentation and explore the section on weighting procedures. From this page, look to the right, and choose the year you are interested in. On the weighting procedures page for the year selected, find the link "Quality Control…" on the right, and from the page that appears, click "Main Quality Control Findings of Interest."

How can I learn more about NAEP sampling?

For more technical information on NAEP sampling, explore NAEP Assessment Sample Design.

NAEP Assessment Sample Design

Each assessment cycle, a sample of students in designated grades within both public and private schools throughout the United States (and sometimes specified territories and possessions) is selected for assessment. In addition, in state assessment years, of which 2011 is an example, the samples of public schools and their students in each state are large enough to support state-level estimates. In all cases, the selection process utilizes a probability sample design in which every school and student has a chance to be selected, and standard errors can be calculated for the derived estimates.

Public School Selection in State Assessment Years

The selection of a sample of public school students for state assessment involves a complex multistage sampling design with the following stages:

- Select public schools within the designated areas,
- Select students in the relevant grades within the designated schools, and
- Allocate selected students to assessment subjects.

The Common Core of Data (CCD) file, a comprehensive list of operating public schools in each jurisdiction that is compiled each school year by the National Center for Education Statistics (NCES), is used as the sampling frame for the selection of sample schools. The CCD also contains information about grades

served, enrollment, and location of each school. In addition to the CCD list, a set of specially sampled [jurisdictions](#) is contacted to determine if there are any newly formed public schools that were not included in the lists used as sampling frames. Considerable effort is expended to increase the survey coverage by locating public schools not included in the most recent CCD file.

As part of the selection process, public schools are combined into groups known as [strata](#) on the basis of various school characteristics related to achievement. These characteristics include the physical location of the school, extent of minority enrollment, state-based achievement scores, and median income of the area in which the school is located. [Stratification](#) of public schools occurs within each state. Grouping schools within strata by such selected characteristics provides a more ordered selection process with improved reliability of the assessment results.

On average, a sample of approximately 100 grade-eligible public schools is selected within each jurisdiction; within each school, about 60 students are selected for assessment. Both of these numbers may vary somewhat, depending on the number and enrollment size of the schools in a jurisdiction, and the scope of the assessment in the particular year. Students are sampled from a roster of individual names, not by whole classrooms. The total number of schools selected is a function of the number of grades to be assessed, the number of subjects to be assessed, and the number of states participating.

Private School Selection in State Assessment Years

In years in which state-level samples are drawn for public schools, private schools are classified by type (e.g., Roman Catholic, Lutheran, etc.), and are grouped for sampling by geography (census division), degree of urbanization of location, and minority enrollment. About 700 private schools, on average, are included, with up to 60 students per school selected for assessment. These samples are not large enough to support state-level estimates for private schools. Thus, inferences for private schools are limited to the national level, even in years when public school assessments are state-specific.

A national sample of private schools in all grades is then drawn from a list compiled through the Private School Universe Survey (PSS), which is a mail survey of all U.S. private schools carried out biennially by the U.S. Census Bureau under contract to NCES. The PSS list is updated for new schools only for a sample of Roman Catholic dioceses.

National-Only Assessment Years

In years when the NAEP samples are intended only to provide representation at the national level and not for each individual state, the public and private school selection process is somewhat different. Rather than selecting schools directly from lists of schools, the first stage of sampling involves selecting a sample of some 50 to 100 geographic primary sampling units (PSUs). Each PSU is composed of one or more counties. They vary in size considerably, and generally about 1,000 PSUs are created in total, from which a sample is selected. Within the set of selected PSUs, public and private school samples are selected using similar procedures to those described above for the direct sampling of schools from lists. The samples are clustered geographically, which results in a more efficient data collection process. The selection of PSUs is not necessary when the sample sizes are large in each state, as in state assessment years.

Last updated 18 July 2017 (GF)

Printer-friendly Version

Achievement Gaps

How Black and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress

Statistical Analysis Report

Achievement Gaps

How Black and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress

Statistical Analysis Report

July 2009

Alan Vanneman
Linda Hamilton
Janet Baldwin Anderson
NAEP Education Statistics Services Institute

Taslima Rahman
National Center for Education Statistics

Executive Summary

In 2007, mathematics scores for both Black and White public school students in grades 4 and 8 nationwide, as measured by the main NAEP assessments of the National Assessment of Educational Progress (NAEP), were higher than in any previous assessment, going back to 1990. This was also true for Black and White fourth-graders on the NAEP 2007 Reading Assessment. For grade 8, reading scores for both Black and White students were higher in 2007 than in the first reading assessment year, 1992, as well as the most recent previous assessment year, 2005.

White students, however, had higher scores than Black students, on average, on all assessments. While the nationwide gaps in 2007 were narrower than in previous assessments at both grades 4 and 8 in mathematics and at grade 4 in reading, White students had average scores at least 26 points higher than Black students in each subject, on a 0-500 scale. This report will use results from both the main NAEP and the long-term trend NAEP assessments to examine the Black-White achievement gaps, and changes in those gaps, at the national and state level.

The main NAEP 2007 Reading and Mathematics Assessments included grade 4 and grade 8 students both nationally and for all 50 states, as well as the Department of Defense Education Activity (DoDEA) and the District of Columbia (hereinafter referred to as states). Not all states had Black (or White) student populations large enough to provide reliable data, and not all states participated in the earliest NAEP state assessments.

Most of the data in this report comes from the main NAEP assessments, supplemented with some data from the NAEP long-term trend assessments. Main NAEP assessments, which began in 1990 for mathematics and 1992 for reading, are administered at the fourth and eighth grades, both nationally and at the state level. Because main NAEP only assesses public schools in its state assessments, this report contains only public school results. The most recent results in this report are for 2007.

NAEP long-term trend assessments are administered by age rather than grade. This report references long-term trend assessment public school results from the earliest assessment through 2004, with results for ages 9 and 13 instead of grades 4 and 8. The long-term trend assessments provide public school results for mathematics going back to 1978 and for reading going back to 1980, at ages 9, 13, and 17, at the national level only, on a 0-500 point scale.

At both ages 9 and 13, mathematics scores for both Black and White students were higher in 2004 than in any previous assessment. The 23-point Black-White achievement gap in mathematics for age 9 public school students in 2004 was narrower than in the first assessment in 1978 but not significantly different from the gap in the most recent previous assessment in 1999. The same was true for the 26-point gap at age 13.

For age 9 reading, scores for both Black and White students were higher in 2004 than in any previous assessment, going back to 1980. The 26-point gap between Black and White students in 2004 was not significantly different from the gap in 1980, but was narrower than the gap in 1999. At age 13 reading, scores were higher for Black students in 2004 than in 1980, but did not show a significant difference from 1999. Scores for White students were not significantly different for either comparison year. The 21-point gap in student performance at age 13 reading in 2004 was narrower than in both 1980 and 1999.

The following two sections summarize state-level achievement gaps between Black and White students in the main NAEP assessments in mathematics and reading.

State Black-White Achievement Gaps—Mathematics

- At the state level, gaps in grade 4 mathematics existed in 2007 in the 46 states for which results were available. In 15 states, the 2007 gaps were narrower than in 1992, as Black students demonstrated a greater gain in average scores than that of the White students.

- At grade 8, mathematics gaps existed in 2007 in the 41 states for which results were available. The gaps were narrower in 2007 than in 1990 in four states: Arkansas, Colorado, Oklahoma, and Texas. In all four, scores for both Black and White students increased, but scores for Black students increased more.

- At grade 4, five states had mathematics gaps in 2007 that were larger than the national gap of 26 points, while 10 states had gaps that were smaller.

- At grade 8, seven states had mathematics gaps in 2007 that were larger than the national gap of 31 points, while 12 had gaps that were smaller.

State Black-White Achievement Gaps—Reading

- At the state level, gaps in grade 4 reading existed in 2007 in the 44 states for which results were available. Gaps narrowed from 1992 to 2007 in Delaware, Florida, and New Jersey, due to larger increases in Black students' scores.

- At grade 8, reading gaps existed in 2007 in 41 of the 42 states for which results were available. In Hawaii, the 7-point difference between Black and White students' scores in 2007 was not statistically significant, and thus there was no gap for Hawaii. There was no significant change in the gap in any state from 1998 to 2007.

- At grade 4, eight states had reading gaps that were larger than the 2007 national gap of 27 points, while nine had gaps that were smaller.

- At grade 8, one state had a reading gap that was larger than the 2007 national gap of 26 points, while nine had gaps that were smaller.

The NAEP reading and mathematics scales make it possible to examine relationships between students' performance and various background factors measured by NAEP, such as race. However, a relationship that exists between achievement and another variable does not reveal its underlying cause, which may be influenced by a number of other variables. Similarly, the assessments do not reflect the influence of unmeasured variables. At the state level, changes in the size of the achievement gap between Black and White students could be affected by demographic changes in the size and makeup of the populations involved, as well as policy changes in the schools and communities. The results of this study are most useful when they are considered in combination with other knowledge about the student population and the education system, such as trends in instruction, changes in the school-age population, and societal demands and expectations.

This report focuses on the size of the achievement gap between Black and White students and the direction of average scores within states, regardless of the states' scores. Large gaps may occur in some states with scores above the national average, as well as in states with scores below the national average. Similarly, small gaps may occur in states with scores above or below the national average. All differences discussed in this report are statistically significant at the .05 level after controlling for multiple comparisons. The technical notes for this report provide information about sampling, accommodations, interpreting statistical significance, and other technical features. For more information on both the main NAEP and long-term trend assessments, see appendix A.

Contents

List of Tables

List of Figures

Introduction

The past half century has witnessed considerable gains in educational attainment in the United States. Between 1950 and 2005, the percentage of young adults ages 25-29 who had completed high school rose from 53 to 86. For White young adults, the percentage increased from 56 to 93, and for Black young adults it increased from 24 to 86.[1]

There have also been gains in educational achievement. National and state mathematics scores in grades 4 and 8 on the National Assessment of Educational Progress (NAEP) were at their highest levels in 2007.[2] Reading scores for the nation and a substantial number of states have also increased since the early 1990s.[3]

Although scores have increased for both Black students and White students, on average Black students do not perform as well as their White peers. At the national level, the fourth-grade Black-White achievement gap in mathematics for 2007 was narrower than in 1990, while the fourth-grade reading gap was narrower than in either 1992 or 2005. At the eighth grade, the gap in mathematics was narrower in 2007 than in 2005, while the reading gap did not change significantly compared to either prior assessment year.

The Elementary and Secondary Education Act[4] when first authorized intended to improve the educational achievement of low-performing students, particularly low-income students and Black students. Subsequent reauthorizations of the act have reaffirmed the importance of closing the achievement gaps. This report uses NAEP data to examine the progress of the nation and each of the states in reducing the gap between Black and White students at grades 4 and 8 in both reading and mathematics.

Issues relating to the Black-White achievement gap have been addressed by a number of recent studies. *Status and Trends in the Education of Racial and Ethnic Minorities*,[5] issued by the National Center for Education Statistics (NCES), for example, examined the education of all major racial and ethnic groups in the United States from pre-kindergarten through the postsecondary level, along with employment and income data for these groups. The report identified a variety of factors which are correlated with the achievement gap between Black and White students. For example, Black students were more likely than White students to come from families living in poverty, which is associated with lower educational performance.

Other reports have used NAEP data in analyses attempting to isolate important factors related to the Black-White achievement gap. For example, *The Family: America's Smallest School*,[6] issued by the Educational Testing Service, correlates student achievement, as measured by NAEP, with four home factors: the presence of two parents in the home, the hours children spend watching television, the hours parents spend reading to them, and the frequency of absence from school. Compared to White students, Black children were less likely to come from a family with both parents in the home, spent more hours watching television, were read to by their parents for fewer hours, and were more likely to be absent from school.

Another report issued by the Educational Testing Service, *Parsing the Achievement Gap II*,[7] considered 16 factors previously identified as being correlated with how well students performed in school. Seven were school-related (including, for example, curriculum rigor and teacher preparation), eight "before and after" school factors (including, for example, weight at birth, exposure to lead, and excessive TV watching), and the "home school connection," parent

---

[1] Snyder, T.D., Dillow, S.A., and Hoffman, C.M. (2007). *Digest of Education Statistics 2006* (NCES 2007-017). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

[2] Lee, J., Grigg, W., and Dion, G. (2007). *The Nation's Report Card: Mathematics 2007* (NCES 2007-494). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, D.C.

[3] Lee, J., Grigg, W., and Donahue, P. (2007). *The Nation's Report Card: Reading 2007* (NCES 2007-496). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, D.C.

[4] Elementary and Secondary Education Act of 1965, P.L. 89-10, 79 Stat. 27.

[5] KewalRamani, A., Gilbertson, L., Fox, M., and Provasnik, S. (2007). *Status and Trends in the Education of Racial and Ethnic Minorities* (NCES 2007-039). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

[6] Barton, P., and Coley, R. (2007). *The Family: America's Smallest School.* Princeton, NJ: Educational Testing Service.

[7] Barton, P., and Coley, R. (2009). *Parsing the Achievement Gap II.* Princeton, NJ: Educational Testing Service.

participation. Using data from NAEP and other sources, the report said that for all 16 factors there were gaps that favored White students over Black students—for example, White students were more likely than Black students to attend schools offering rigorous curriculums and less likely to suffer from low birth weight.

This report uses data from both the "main NAEP" and the NAEP long-term trend assessments. NCES and the National Assessment Governing Board, which sets policy for NAEP, have maintained comparability of data for both main and long-term trend NAEP. Main NAEP assessments, which began in 1990 for mathematics and 1992 for reading, are administered at the fourth and eighth grades, both nationally and at the state level. The most recent administration was in 2007. The long-term trend assessments provide public school results for mathematics going back to 1978 and for reading going back to 1980, at ages 9, 13, and 17, at the national level only. The most recent long-term trend report available at the time of the preparation of this report contains results for the assessments administered in 2004. Discussion of main NAEP grade 12 assessments is omitted in this report because these assessments are conducted at the national level only.

While the main NAEP assessments do not go as far back in time as the long-term trend assessments, they allow the examination of trends in the Black and White performance gap in every state, plus the District of Columbia and the U.S. Department of Defense Education Activity (DoDEA) schools. In addition, the main NAEP assessments use frameworks that are more closely aligned with current practices regarding instructional content; they include more questions overall and more questions that require a written response; and they employ much larger samples than long-term NAEP.

All data presented in this report for main NAEP are for public school students only. Main NAEP and long-term trend provide national results for both public and private school students, but NAEP state results are for public school students only. To maintain consistency of data for comparison purposes, this report uses only public school data at the national level as well.

The major questions addressed in this study are: 1) how do gaps in 2007 compare to the gaps in the initial and most recent prior years of the NAEP national and state assessment series? And 2) how do states compare to the nation in 2007? The current report presents these results in graphs that show the NAEP achievement gaps in a format that makes it possible to see at a glance the national and state gaps results for all available years.

In previous NAEP reports, achievement gaps results have been available to users in two ways: 1) online, using the NAEP Data Explorer, and 2) by year, in the report cards for a given assessment. The NAEP Achievement Gaps report is the first NCES publication to present the Black and White NAEP achievement gaps across time for all the states and the nation, including results for every assessment year since state assessments began.

States first participated in the eighth-grade mathematics assessment in 1990, the fourth-grade reading and mathematics assessments in 1992, and the eighth-grade reading assessment in 1998. The No Child Left Behind Act of 2001 requires each state, beginning in 2003, to participate in the NAEP mathematics and reading assessments if they are to receive Title I education funding (Public Law 107-110 Title I Part A, Sec. 1111). Prior to the passage of the Act, participation was voluntary and about 40 states participated in each assessment. (In this report, "state" and "jurisdiction" will be used interchangeably to refer to the 50 states, the District of Columbia, and the Department of Defense Education Activity (DoDEA) schools.) Additional information about the years when the national and state assessments were administered is in appendix B.

Sources of the Main NAEP data

This report presents national data from the NAEP reading and mathematics assessments for Black and White public school students at the fourth and eighth grades. Only results for White (non-Hispanic) and Black (non-Hispanic) public school students are contained in this report. Additional information on the national and state assessments is given in appendix B.

Administration of main NAEP national and state reading and mathematics assessments

			1990	1992	1994	1996	1998	2000	2002	2003	2005	2007
Reading	**4th Grade**	**National**		✓	✓		✓	✓	✓	✓	✓	✓
		State		✓	✓		✓		✓	✓	✓	✓
	8th Grade	**National**		✓	✓		✓	✓	✓	✓	✓	✓
		State					✓		✓	✓	✓	✓
Mathematics	**4th Grade**	**National**	✓	✓		✓		✓		✓	✓	✓
		State		✓		✓		✓		✓	✓	✓
	8th Grade	**National**	✓	✓		✓		✓		✓	✓	✓
		State	✓	✓		✓		✓		✓	✓	✓

In 2007, Black and White students together comprised about three-fourths of the nation's public school students at the fourth and eighth grades. At the fourth-grade level, 58 percent of assessed students were White and 16 percent were Black. At the eighth-grade level, 60 percent of assessed students were White and 16 percent were Black.

In the earliest main NAEP assessments, students with disabilities and English language learners did not receive accommodations. Since 1994 (1996 at the state level) students receiving accommodations on their state assessment received the same accommodations on NAEP, as long as NAEP approved them (see appendix A for details.)

In 2007, the reading assessment was given to 183,000 fourth-graders and 155,000 eighth-graders, while the mathematics assessment was given to 190,000 fourth-graders and 147,000 eighth-graders. The main NAEP samples are so large because they include representative samples for each of the 50 states, plus the District of Columbia and Department of Defense school system for Armed Forces dependents in the United States and overseas. This allows examination of the achievement gaps for public school students for individual states as well as for the nation as a whole.

NAEP assessments are conducted in a six-week window starting in January of each assessment year. The same assessment is administered in both the national and state assessments. Because the content of the assessments given to fourth-graders and eighth-graders differs, scores for the two grades should be compared with caution, even though the scores appear on similar 0-500 scales. Scores for reading and mathematics cannot be compared because the two assessments are scaled independently. See appendix A for more details.

Sources of the Long-Term Trend NAEP data

This report presents national data for public school students aged 9 and 13 from the 1978, 1982, 1986, 1990, 1992, 1994, 1996, 1999, and 2004 mathematics long-term trend assessments and the 1982, 1986, 1990, 1992, 1994, 1996, 1999, and 2004 reading long-term trend assessments. Unlike the main NAEP assessments, these assessments did not allow accommodations for students with disabilities and English language learners for the years included in this report. Sample sizes for the 2004 long-term trend assessments were 7,500 (9-year-old students) and 8,300 (13-year-old students) for reading and 7,300 (9-year-old students) and 7,500 (13-year-old students) for mathematics. See appendix A for more details.

Understanding score gaps

Ways the gap can change

The achievement gap between Black and White students is defined as the difference between the average score for Black students and the average score for White students. Comparisons are made for main NAEP between the most recent assessment year (2007) and all previous assessment years. Only changes between the earliest assessment year and 2007, and between 2005 and 2007, are discussed. For long-term trend, only changes between the earliest assessment year and 2004, and between 1999 and 2004, are discussed.

Changes in the size of the achievement gap depend on both changes in the average scores for Black and White students and the rate of change in those scores. Generally, widening gaps are seen as undesirable, while narrowing gaps are seen as desirable. However, it is possible for the gap to widen even if scores for both Black students and White students increase, if scores for the higher scoring group increase more than scores for the other group. And it is also possible for the gap to narrow even if scores for both Black and White students decline, if scores for the higher scoring group decline more than those of the other group. The following images illustrate the various ways that gaps can narrow.

Ways gaps can narrow

The average scores of *both* groups increase, while the score of the lower performing group increases even more.

The average score of the *higher* performing group does not change, while the score of the lower performing group increases.

The average score of the *higher* performing group declines, while the score of the lower performing group increases.

The average score of the *higher* performing group declines, while the score of the lower performing group does not change.

The average scores of *both* groups decline, but the score of the higher performing group declines even more.

It is important to note that although NAEP data can identify gaps and changes in gaps, these data cannot explain why gaps exist or why they change. NAEP assessments are designed to measure student performance and identify factors associated with it, not to identify or explain the causes of differences in student performance.

Understanding statistical significance

NAEP data are based on samples of students, and the results are subject to sampling and measurement error. Statistical tests are used to determine whether the differences between average scores are statistically significant—that is, whether they exceed the margin of error. Changes in average scores for Black students and White students and changes in the size of the gap between these scores are analyzed separately. Therefore, it is possible for the size of the achievement gap to increase or decrease even though the average scores of neither Black nor White students changed statistically significantly during the same period.

The term "significant" is not intended to imply a judgment about the absolute magnitude or the educational relevance of the differences. It is intended to identify statistically reliable population differences to help inform discussion among policymakers, educators, researchers, and the public.

Beginning in 2002, the main NAEP national sample was obtained by aggregating the samples from each state, rather than by using an independently selected national sample. As a result, the national samples in mathematics and reading were larger in 2003, 2005, and 2007 than in previous assessment years. Thus, smaller score differences between years or between student groups were found to be statistically significant than would have been detected in previous assessments. All differences discussed in the text are significant at the .05 level with appropriate adjustments for part-to-whole and multiple comparisons.

Statistical comparisons of NAEP scores from different assessment years are made using a multiple comparison procedure (see appendix A, "Conducting multiple tests," for details). However, in figures 9, 11, 21, and 23, comparisons of the size of the Black-White achievement gap

4

for each state to the national gap are made using pairwise comparisons, where each state is compared to the nation one at a time. For this reason, the results shown in these four figures may not correspond to results obtained from the NAEP Online Data Tool, which currently does not permit pairwise comparisons for this type of gap analysis.

Cautions in interpreting the data

All results given here are in terms of average scores, which reflect a wide range of student performance. Many Black students score above the average for White students and many White students score below the average for Black students. For detailed information on variations in performance, including standard deviations, consult the NAEP Data Explorer online at http://nces.ed.gov/nationsreportcard/nde/viewresults.asp

The analysis of NAEP data contained in this report should not be seen to imply causal relations. Simple cross-tabulations of a variable with measures of educational achievement, like the ones presented here, cannot be considered as evidence that differences in the variable cause differences in educational achievement. As noted earlier, NAEP surveys are not designed to identify causal relationships. There are many possible reasons why the performance of one group of students will differ from that of another. Inferences related to student group performance should take into consideration the many socioeconomic and educational factors that may also be associated with performance.

All statistical tests are performed using unrounded scale scores. The Black-White achievement gap is calculated by subtracting the average scale score for Black students from the average scale score for White students. Because all results are presented as rounded numbers, occasionally the lower scale score plus the gap will not equal the higher scale score shown in this report's graphics.

How this report is organized

The remainder of this report presents first mathematics and then reading results. In each section, long-term trend results are presented first, giving national results only for public school students ages 9 and 13. These are followed by both national and state results for public school fourth- and eighth-graders from main NAEP. National data from main NAEP are also presented by 1) gender and 2) eligibility categories for the National School Lunch Program. The last section consists of an appendix that contains relevant technical notes and supplemental tables.

Long-Term Trend Results
for Black and White 9- and 13-Year-Olds

Trends in mathematics scores and achievement gaps, 1978–2004

Mathematics scores for both 9- and 13-year-old Black and White students were higher in 2004 than on any previous long-term trend assessment (figures 1 & 2). In addition, the score gaps for Black and White students were narrower in 2004 than in the first assessment in 1978 for both age groups, as scores of Black students showed a greater increase than those of White students. The gaps in 2004 were not significantly different from the gaps in 1999.

Figure 1. **Trends in average mathematics scale scores and score gaps for White students and Black students at age 9: Various years, 1978–2004**
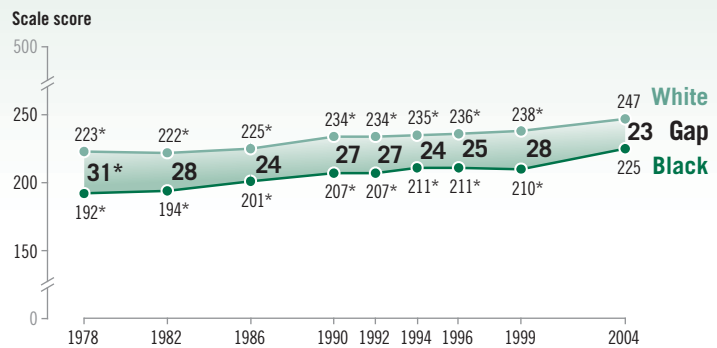
* Significantly different (p<.05) from 2004.
NOTE: Detail may not sum to totals due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1978–2004 Long-Term Trend Mathematics Assessments.

Figure 2. **Trends in average mathematics scale scores and score gaps for White students and Black students at age 13: Various years, 1978–2004**
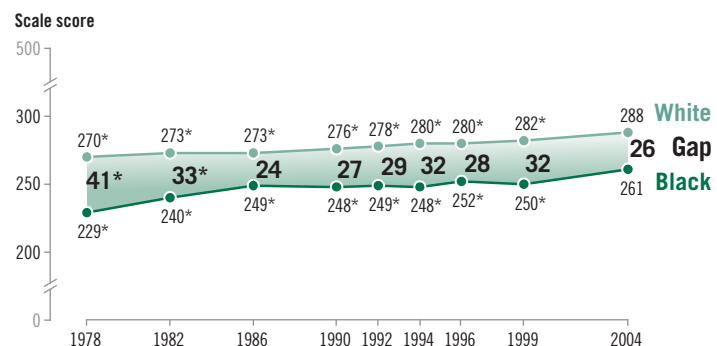
* Significantly different (p<.05) from 2004.
NOTE: Detail may not sum to totals due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1978–2004 Long-Term Trend Mathematics Assessments.

Main NAEP National Results for Black and White Fourth- and Eighth-Graders

Trends in mathematics scores and achievement gaps, 1990–2007

In main NAEP, average fourth-grade mathematics scores for the nation were higher in 2007 than in 1990 for both Black and White public school students (figure 3). The greater increase for Black fourth-graders resulted in the gap narrowing from 31 points in 1990 to 26 points in 2007. From 2005 to 2007, scores increased for both Black and White students, but there was no significant change in the gap.

Average mathematics scores were higher in 2007 than in 1990 for both Black and White eighth-graders (figure 4). The 31-point gap in 2007 was not significantly different from the 33-point gap in 1990. However, the gap was narrower in 2007, at 31 points, than in 2005, at 33 points. Although scores for both groups were higher in 2007, a greater increase in Black students' scores caused the gap to narrow. The 2-point decrease in the gap from 2005 to 2007 was significant while the 2-point decrease from 1990 to 2007 was not. It is possible that the smaller standard errors in 2005, due to the increased sample size in that year, allowed the difference in 2005 to be identified as statistically significant.

Figure 3. Mathematics achievement score gaps between Black and White public school students at grade 4: Various years, 1990–2007
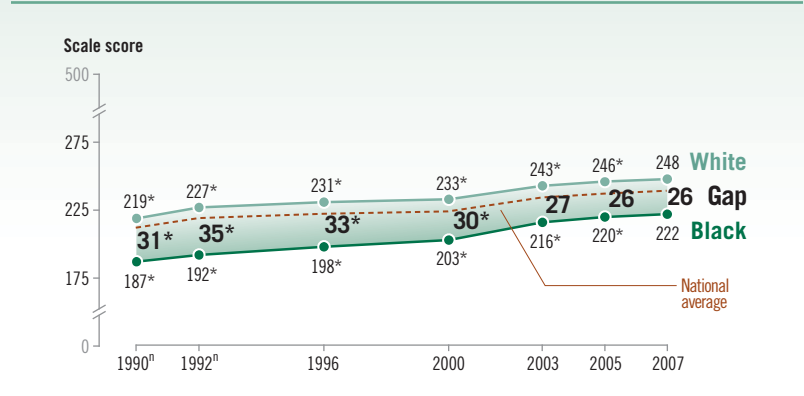
[n] Accommodations were not permitted for this assessment.
* Significantly different (p<.05) from 2007.
NOTE: Detail may not sum to totals due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1990–2007 Mathematics Assessments.

Figure 4. Mathematics achievement score gaps between Black and White public school students at grade 8: Various years, 1990–2007

[n] Accommodations were not permitted for this assessment.
* Significantly different (p<.05) from 2007.
NOTE: Detail may not sum to totals due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1990–2007 Mathematics Assessments.

7

Mathematics scores and achievement gaps by gender, 1990–2007

Average mathematics scores were higher in 2007 than in 1990 for the nation's Black and White fourth-graders, regardless of gender (figure 5). Among females, the gap was narrower in 2007 as the average score gains of Black females were greater than those of their White peers. Among fourth-grade males, the Black-White gap did not change significantly.

In addition to the 17-year gain, mathematics scores also increased during the two-year period, 2005 to 2007, for both Black and White fourth-graders, regardless of gender. However, the gaps did not change significantly either for males or for females during this period.

In 2007, average mathematics scores were higher than they had been in 1990 for Black and White eighth-graders (figure 6). However, the Black-White mathematics gap did not change significantly for either males or females.

At grade 8, mathematics scores increased from 2005 to 2007 for Black and White students, regardless of gender (figure 6). Female eighth-graders showed a narrowing of the gap during this period as Black females' scores increased more than those of White females, while the gap for males did not change significantly.

Figure 5. Gaps in average mathematics scores between Black and White public school students at grade 4, by gender: Various years, 1990–2007

Male

Female

Scale score

White
Gap
Black

Male
500
275

225

175

0

219* 228* 232* 236* 244* 247* 249
                                        28
187* 192* 197* 202* 216* 219* 221

1990[n] 1992[n] 1996 2000 2003 2005 2007

Female

218* 225* 230* 231* 241* 244* 246
30* 34* 32* 27 25 24 24
188* 192* 198* 204* 216* 220* 223

1990[n] 1992[n] 1996 2000 2003 2005 2007
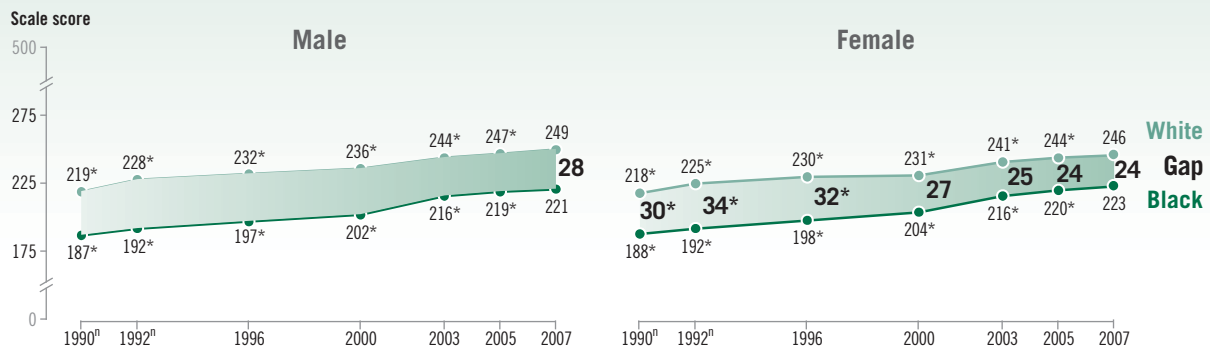
[n] Accommodations were not permitted for this assessment.
* Significantly different (p<.05) from 2007.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1990–2007 Mathematics Assessments.

Figure 6. Gaps in average mathematics scores between Black and White public school students at grade 8, by gender: Various years, 1990–2007

Scale score

White
Gap
Black

Male
500
325

275

225

0

270* 275* 280* 284* 287* 288* 292
35 39* 39 42* 36* 34 34
235* 236* 240* 242* 251* 254* 258

1990[n] 1992[n] 1996 2000 2003 2005 2007

Female

268* 276* 279* 282* 286* 287* 289
31 40* 41* 37* 33* 32* 29
237* 236* 238* 244* 252* 255* 260

1990[n] 1992[n] 1996 2000 2003 2005 2007

[n] Accommodations were not permitted for this assessment.
* Significantly different (p<.05) from 2007.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1990–2007 Mathematics Assessments.

Mathematics scores and achievement gaps by family income, 2003–2007

NAEP uses student eligibility for free or reduced-price school lunch as an indicator of family income. At grade 4, mathematics scores were higher in 2007 than in 2003 and 2005 for all Black and White public school students, regardless of school-lunch eligibility (figure 7). Despite these increases, the only significant Black-White gap change was between 2003 and 2007, for students eligible for reduced-price lunch.

At grade 8, mathematics scores were higher in 2007 than in 2003 and 2005 for all Black and White public school students (figure 8). The Black-White score gaps for students eligible for free or reduced-price lunch narrowed in 2007 in comparison to both previous assessments, as scores for eligible Black students showed greater gains than those of their White peers.

Eligibility for free and reduced-price lunch

NAEP collects data on students' eligibility for the National School Lunch Program (NSLP)—sometimes referred to as the free and reduced-price school lunch program—as an indicator of family economic status. Eligibility for free and reduced-price lunch is based on students' family income in relation to the federally established poverty level.

Not eligible: Students who are not eligible for the program because their family's income is above 185 percent of the poverty level.

Eligible for reduced-price lunch: Students who are eligible for reduced-price lunch because their family's income is between 130 percent and 185 percent of the poverty level.

Eligible for free lunch: Students who are eligible for free lunch because their family's income is below 130 percent of the poverty level.

As a result of improvements in the quality of the data on students' eligibility for NSLP, the percentage of students for whom information was not available has decreased in comparison to the percentages reported prior to the 2003 assessment. Therefore, trend comparisons are only made back to 2003 in this report.

Table 1. Percentage of public school students assessed in NAEP mathematics by eligibility for free or reduced-price school lunch, race/ethnicity and grade: 2003, 2005, and 2007

	Not eligible		Eligible for reduced-price lunch		Eligible for free lunch	
	Black	White	Black	White	Black	White
Grade 4						
2007	26	72	7	6	66	21
2005	25	71	8	7	66	20
2003	24	72	9	8	66	19
Grade 8						
2007	32	76	7	5	60	18
2005	31	75	9	6	58	17
2003	32	76	9	6	56	15

NOTE: Detail may not sum to totals due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Mathematics Assessments.

Figure 7. Gaps in average mathematics scores between Black and White public school students at grade 4, by eligibility for free or reduced-price school lunch: 2003, 2005, and 2007
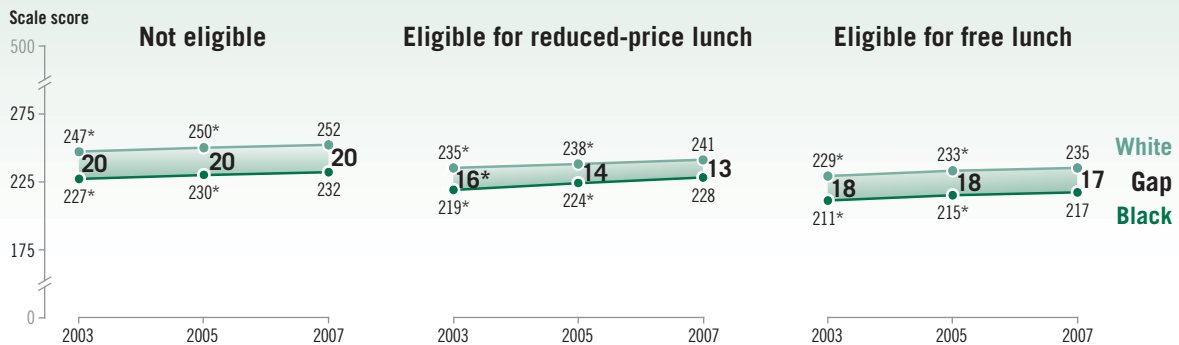
* Significantly different (p<.05) from 2007.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Mathematics Assessments.

Figure 8. Gaps in average mathematics scores between Black and White public school students at grade 8, by eligibility for free or reduced-price school lunch: 2003, 2005, and 2007
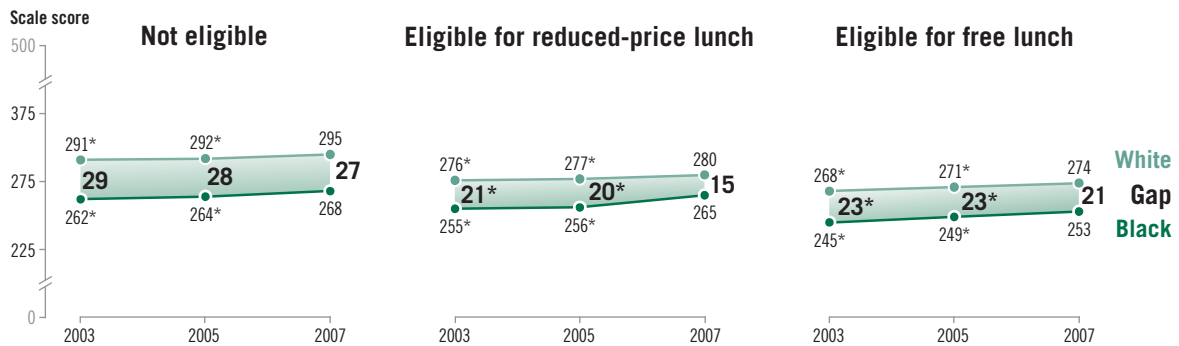
* Significantly different (p<.05) from 2007.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Mathematics Assessments.

Main NAEP State Results for Black and White Fourth- and Eighth-Graders

The NAEP state mathematics assessments were administered to public school fourth-graders in 1992, 1996, 2000, 2003, 2005, and 2007 and to public school eighth-graders in 1990, 1992, 1996, 2000, 2003, 2005, and 2007. Before 2003, states were not required to participate in NAEP to qualify for Title I education funds. Typically, 40 or more states participated in each prior assessment. In 2003, 2005, and 2007, all 50 states, the District of Columbia, and the DoDEA participated.

State results are presented in two ways. Comparisons of fourth-grade mathematics gaps in 2007 between each state and the nation are presented in figure 9.

Comparisons of the mathematics gaps within a state over time are presented in a series of small graphs in figure 10. At the top left of each two-page spread, the mathematics scores and gaps for the nation are presented for reference. Each state figure, as well as the national figure, also contains a dotted red line representing the national average for public school students. The data for the national averages are located in the appendix in Table B-2.

State and national mathematics achievement gaps at grade 4, 2007

Ten states had a smaller Black-White gap than the nation's 26-point gap in 2007 (Alaska, Delaware, DoDEA, Hawaii, Kentucky, Louisiana, Mississippi, Oklahoma, Texas, and West Virginia) and five had a gap that was larger (Connecticut, District of Columbia, Illinois, Nebraska, and Wisconsin). In 31 states, the gap was not significantly different from the nation's gap. Gaps that are different from the nation's gap are indicated with an asterisk (figure 9).

The fourth-grade mathematics gap in 2007 was statistically significant in all 46 states for which data could be reported. The gaps ranged from 14 points in Hawaii and West Virginia to 54 points in the District of Columbia.

Figure 9. The Black-White achievement score gap in mathematics for public school students at grade 4, by state or jurisdiction: 2007

Jurisdictions	Black	Gap	White
Nation (public)	222	**26**	248
Alabama	213	**25**	238
Alaska	227	**20***	247
Arizona	219	**28**	246
Arkansas	217	**28**	245
California	218	**29**	247
Colorado	224	**26**	249
Connecticut	220	**32***	252
Delaware	230	**20***	249
District of Columbia	209	**54***	262
DoDEA[1]	227	**19***	246
Florida	225	**25**	250
Georgia	222	**24**	246
Hawaii	230	**14***	244
Illinois	216	**32***	248
Indiana	224	**25**	249
Iowa	224	**21**	245
Kansas	226	**26**	252
Kentucky	219	**19***	238
Louisiana	219	**21***	240
Maine	221	**22**	243
Maryland	223	**29**	251
Massachusetts	232	**25**	257
Michigan	216	**28**	244
Minnesota	222	**31**	252
Mississippi	217	**22***	239
Missouri	218	**26**	246
Nebraska	211	**33***	244
Nevada	219	**23**	243
New Hampshire	226	**24**	250
New Jersey	232	**23**	255
New Mexico	220	**22**	242
New York	225	**26**	251
North Carolina	224	**27**	251
Ohio	225	**25**	250
Oklahoma	220	**22***	242
Oregon	219	**22**	241
Pennsylvania	222	**26**	249
Rhode Island	219	**23**	242
South Carolina	221	**26**	248
South Dakota	221	**24**	245
Tennessee	214	**26**	240
Texas	230	**23***	253
Virginia	228	**23**	251
Washington	222	**26**	248
West Virginia	223	**14***	237
Wisconsin	212	**38***	250

Scale score

* Significantly different (p<.05) from the nation (public) when comparing one state to the nation at a time.
[1] Department of Defense Education Activity (overseas and domestic schools).
NOTE: States whose Black student population size was insufficient for comparison are omitted. Reporting standards not met for Idaho, Montana, North Dakota, Utah, Vermont, and Wyoming.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessment.

Trends in state mathematics achievement gaps at grade 4, 1992–2007

The Black-White mathematics gap among the nation's public school fourth-graders was narrower in 2007 than in 1992, as Black students' scores showed a greater gain than White students' scores (figure 10, National results). From 2005 to 2007, there was no significant change in the gap.

In 35 states, both Black students and White students achieved higher average scores in mathematics from 1992 to 2007. Fifteen of these states also narrowed the achievement gap as Black students' scores increased more than White students' scores.

Short-term changes were also notable. In Illinois, New Jersey, and Virginia, average scores for both Black and White students increased between 2005 and 2007.

Narrowing of the Gap

In the following 15 states, the gap narrowed between **1992 and 2007** as gains of Black students outpaced the gains of White students.

California	**Michigan**
Connecticut	**Mississippi**
Delaware	**New Jersey**
District of Columbia	**Pennsylvania**
Florida	**South Carolina**
Georgia	**Texas**
Louisiana	**Virginia**
Massachusetts	

In **Rhode Island**, the gap narrowed between **2005 and 2007** as Black students' scores increased while those of White students did not change significantly.

Figure 10.   Gaps in average mathematics scores between Black and White public school students at grade 4, by state: Various years, 1992–2007

Scale score

NATIONAL[1]
(Black: 17%, White: 55%)[2]

White: 227* 230* 233* 243* 246* 248
Gap: 35* 31 30* 27 26 26
Black: 192* 199* 203* 216* 220* 222

1992[n] 1996[n] 2000 2003 2005 2007

Alabama
(Black: 37%, White: 58%)

218* 221* 227* 232* 235* 238
30 29 24 24 24 25
188* 193* 203* 208* 211 213

1992[n] 1996[n] 2000 2003 2005 2007

Alaska
(Black: 5%, White: 55%)

232* 242* 244 247 **White**
25 20 18 **20 Gap**
206* 221 226 227 **Black**

National average

1996[n] 2003 2005 2007

Arizona
(Black: 5%, White: 43%)

225* 228* 230* 241* 243* 246
26 31 22 26 25 28
199* 197* 207* 215 217 219

1992[n] 1996[n] 2000 2003 2005 2007

Arkansas
(Black: 22%, White: 67%)

217* 223* 225* 237* 242* 245
29 30 31 31 29 28
188* 193* 194* 206* 214 217

1992[n] 1996[n] 2000 2003 2005 2007

California
(Black: 7%, White: 27%)

221* 223* 228* 243* 245 247 **White**
39* 34 33 30 29 **29 Gap**
182* 188* 194* 213* 215 218 **Black**

National average

1992[n] 1996[n] 2000 2003 2005 2007

Colorado
(Black: 6%, White: 60%)

227* 232* 243* 247 249
28 35 26 25 26
199* 196* 217 222 224

1992[n] 1996[n] 2003 2005 2007

Connecticut
(Black: 13%, White: 64%)

235* 240* 242* 250 250 252
40* 35 32 32 32 32
195* 205* 210* 217 219 220

1992[n] 1996[n] 2000 2003 2005 2007

Delaware
(Black: 33%, White: 54%)

226* 225* 244* 249 249 **White**
30* 31* 22 23 **20 Gap**
197* 194* 223* 226 230 **Black**

National average

1992[n] 1996[n] 2003 2005 2007

District of Columbia
(Black: 84%, White: 6%)

251* 248 254 262 266 262
62* 65 66* 60 58 **54**
189* 183* 188* 202* 207 209

1992[n] 1996[n] 2000 2003 2005 2007

DoDEA[3]
(Black: 17%, White: 51%)

231* 233* 242* 245 246
22 20 15* 19 19
209* 213* 227 227 227

1996[n] 2000 2003 2005 2007

Florida
(Black: 21%, White: 48%)

224* 227* 243* 247* 250 **White**
34* 33* 28 23 **25 Gap**
189* 193* 215* 224 225 **Black**

National average

1992[n] 1996[n] 2003 2005 2007

See notes at end of figure.

15

Figure 10. Gaps in average mathematics scores between Black and White public school students at grade 4, by state: Various years, 1992–2007—Continued

See notes at end of figure.

Figure 10.   Gaps in average mathematics scores between Black and White public school students at grade 4, by state: Various years, 1992–2007—Continued

Scale score

Massachusetts
(Black: 7%, White: 75%)

White: 231*, 232*, 239*, 247*, 252*, 257
Gap: 36*, 26, 27, 26, 24, 25
Black: 195*, 206*, 213*, 222*, 228, 232

Years: 1992[n], 1996[n], 2000, 2003, 2005, 2007

Michigan
(Black: 21%, White: 71%)

White: 227*, 232*, 237*, 244, 245, 244
Gap: 42*, 34, 37*, 35, 35, 28
Black: 185*, 198*, 199*, 209, 211, 216

Years: 1992[n], 1996[n], 2000, 2003, 2005, 2007

Minnesota
(Black: 8%, White: 78%)

White: 231*, 235*, 238*, 246*, 251, 252
Gap: 38, 40, 30, 28, 32, 31
Black: 193*, 196*, 208*, 219, 219, 222
National average

Years: 1992[n], 1996[n], 2000, 2003, 2005, 2007

Mississippi
(Black: 52%, White: 45%)

White: 219*, 221*, 222*, 236, 238, 239
Gap: 30*, 25, 24, 24, 23, 22
Black: 189*, 196*, 198*, 212*, 216, 217

Years: 1992[n], 1996[n], 2000, 2003, 2005, 2007

Missouri
(Black: 19%, White: 77%)

White: 227*, 230*, 233*, 240*, 240*, 245
Gap: 32, 29, 31, 24, 25, 26
Black: 195*, 200*, 202*, 216, 215, 218

Years: 1992[n], 1996[n], 2000, 2003, 2005, 2007

Montana
(Black: 1%, White: 83%)
Gap data not available

White: 231*, 231*, 238*, 243*, 247
Black
National average

Years: 1996[n], 2000, 2003, 2005, 2007

Nebraska
(Black: 7%, White: 75%)

White: 228*, 231*, 230*, 241*, 244, 244
Gap: 38, 34, 38, 31, 34, 33
Black: 191*, 197*, 193*, 211, 211, 211

Years: 1992[n], 1996[n], 2000, 2003, 2005, 2007

Nevada
(Black: 8%, White: 43%)

White: 224*, 226*, 236*, 240, 243
Gap: 29, 23, 21, 26, 23
Black: 195*, 203*, 215, 214, 219

Years: 1996[n], 2000, 2003, 2005, 2007

New Hampshire
(Black: 2%, White: 91%)

White: 230*, 244*, 246*, 250
Gap: 24
Black: 226
National average

Years: 1992[n], 2003, 2005, 2007

New Jersey
(Black: 14%, White: 57%)

White: 236*, 239*, 248*, 251*, 255
Gap: 38*, 35*, 31*, 27, 23
Black: 198*, 204*, 217*, 224*, 232

Years: 1992[n], 1996[n], 2003, 2005, 2007

New Mexico
(Black: 3%, White: 29%)

White: 224*, 227*, 227*, 237*, 238*, 242
Gap: 22, 20, 25, 22
Black: 202*, 216, 213, 220

Years: 1992[n], 1996[n], 2000, 2003, 2005, 2007

New York
(Black: 19%, White: 53%)

White: 228*, 233*, 238*, 246*, 247*, 251
Gap: 31, 31, 27, 26, 25, 26
Black: 197*, 202*, 210*, 219, 222, 225
National average

Years: 1992[n], 1996[n], 2000, 2003, 2005, 2007

See notes at end of figure.

Figure 10. Gaps in average mathematics scores between Black and White public school students at grade 4, by state: Various years, 1992–2007—Continued
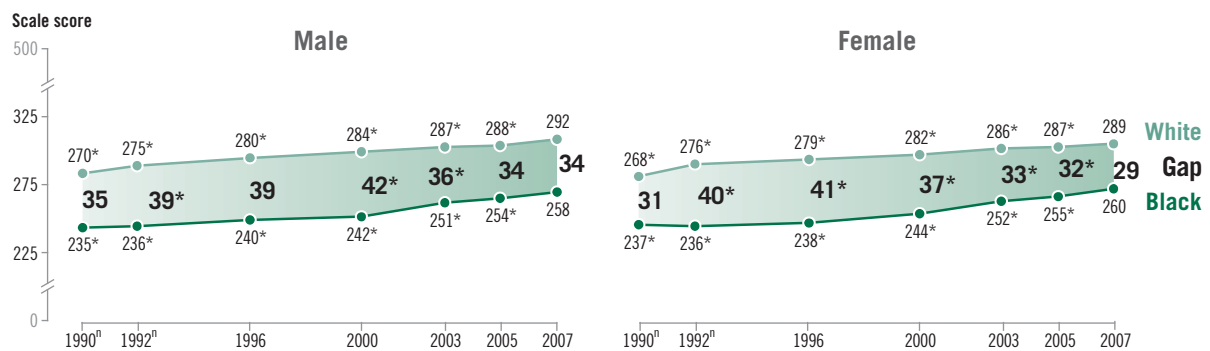

NATIONAL[1]
(Black: 17%, White: 55%)[2]

North Carolina
(Black: 28%, White: 55%)

North Dakota
(Black: 2%, White: 87%)
Gap data not available

Ohio
(Black: 18%, White: 75%)

Oklahoma
(Black: 11%, White: 58%)

Oregon
(Black: 3%, White: 71%)

Pennsylvania
(Black: 14%, White: 77%)

Rhode Island
(Black: 8%, White: 70%)

South Carolina
(Black: 36%, White: 57%)

South Dakota
(Black: 2%, White: 83%)

Tennessee
(Black: 26%, White: 69%)

Texas
(Black: 15%, White: 36%)

See notes at end of figure.

18

Figure 10.   Gaps in average mathematics scores between Black and White public school students at grade 4, by state: Various years, 1992–2007—Continued

Scale score

Utah
(Black: 1%, White: 80%)

Gap data not available

225* 228* 230* 238* 242 244

1992[n] 1996[n] 2000 2003 2005 2007

Vermont
(Black: 2%, White: 94%)

Gap data not available

225* 232* 242* 244* 247

1996[n] 2000 2003 2005 2007

Virginia
(Black: 26%, White: 58%)

228* 230* 237* 246* 247* 251 **White**
29* **27** **26** **23** **24** **23 Gap**
199* 203* 211* 223* 224* 228 **Black**

National average

1992[n] 1996[n] 2000 2003 2005 2007

Washington
(Black: 6%, White: 65%)

229* 242* 246 248
27 **19** **15*** **26**
202* 222 231* 222

National average

1996[n] 2003 2005 2007

West Virginia
(Black: 5%, White: 93%)

216* 224* 224* 231* 231* 237
15 **19** **20** **10** **5*** **14**
201* 205* 205* 221 226 223

1992[n] 1996[n] 2000 2003 2005 2007

Wisconsin
(Black: 10%, White: 77%)

233* 236* 243* 247* 250 **White**
37 **38** **35** **37** **38 Gap**
195* 198* 209 210 212 **Black**

National average

1992[n] 1996[n] 2000 2003 2005 2007

Wyoming
(Black: 2%, White: 84%)

Gap data not available

227* 225* 231* 243* 245 246 **White**

National average

1992[n] 1996[n] 2000 2003 2005 2007

[n] Accommodations were not permitted for this assessment.
* Significantly different (p<.05) from 2007.
[1] National results for assessments prior to 2002 are based on the national sample, not on aggregated state samples.
[2] Black and White percentages are based on students tested in 2007.
[3] Department of Defense Education Activity (overseas and domestic schools). Before 2005, DoDEA overseas and domestic schools were separate jurisdictions in NAEP. Pre-2005 data presented here were recalculated for comparability.
NOTE: Detail may not sum to totals due to rounding. Where data are not present, the jurisdiction did not participate or did not meet the minimum participation guidelines for reporting. State-level data were not collected in 1990. Comparative performance results may be affected by changes in exclusion rates for students with disabilities and English language learners in the NAEP samples.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1992–2007 Mathematics Assessments.
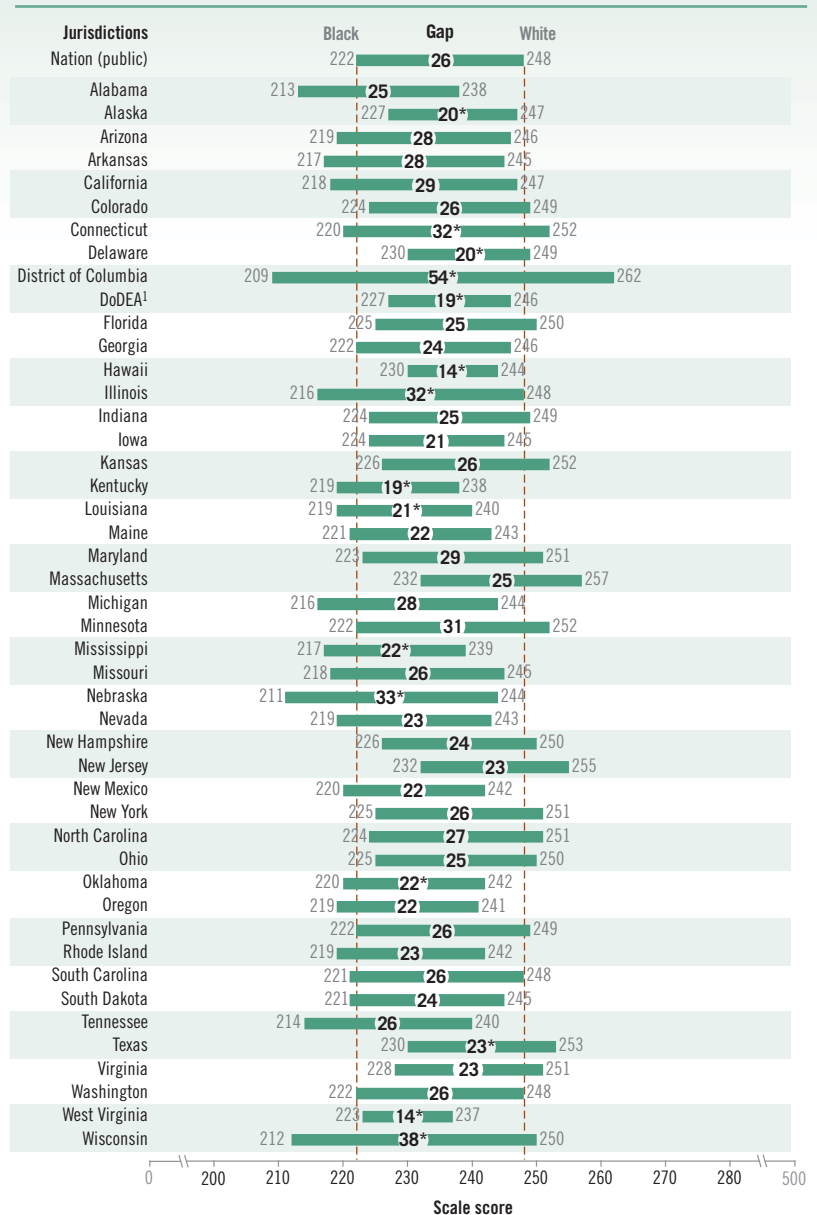
State and national mathematics achievement gaps at grade 8, 2007

Twelve states had a smaller gap than the nation's 31-point gap in 2007 (Alaska, Arizona, Colorado, DoDEA, Georgia, Kentucky, Louisiana, Mississippi, New Mexico, Oklahoma, Oregon, and South Carolina) and seven had a gap that was larger (Connecticut, Illinois, Maryland, Massachusetts, Michigan, Nebraska, and Wisconsin). In 22 states, the gap was not significantly different from the nation's gap. Gaps that are different from the nation's gap are indicated with an asterisk (figure 11).

The eighth-grade mathematics gap in 2007 was statistically significant in all 41 states for which data could be reported. The gaps ranged from 16 points in Oregon to 51 points in Nebraska.

Figure 11. The Black-White achievement score gap in mathematics for public school students at grade 8, by state or jurisdiction: 2007

Jurisdictions	Black	Gap	White
Nation (public)	259	(31)	290
Alabama	246	(32)	278
Alaska	271	(23*)	294
Arizona	266	(23*)	289
Arkansas	254	(28)	282
California	253	(35)	287
Colorado	272	(24*)	296
Connecticut	255	(38*)	293
Delaware	265	(29)	294
DoDEA[1]	272	(19*)	291
Florida	259	(29)	289
Georgia	261	(27*)	288
Illinois	253	(38*)	291
Indiana	259	(32)	290
Iowa	257	(31)	288
Kansas	267	(28)	295
Kentucky	257	(25*)	282
Louisiana	258	(25*)	283
Maryland	265	(36*)	300
Massachusetts	264	(40*)	305
Michigan	244	(41*)	285
Minnesota	260	(37)	297
Mississippi	251	(28*)	279
Missouri	253	(34)	288
Nebraska	240	(51*)	291
Nevada	255	(26)	282
New Jersey	264	(35)	298
New Mexico	264	(21*)	285
New York	258	(32)	290
North Carolina	266	(29)	295
Ohio	258	(33)	291
Oklahoma	258	(22*)	280
Oregon	272	(16*)	289
Pennsylvania	257	(36)	293
Rhode Island	250	(34)	284
South Carolina	265	(27*)	293
Tennessee	254	(28)	282
Texas	271	(29)	300
Virginia	268	(28)	296
Washington	264	(26)	291
West Virginia	250	(21)	271
Wisconsin	247	(45*)	292

0   240   250   260   270   280   290   300   310   320   500

Scale score

* Significantly different (p<.05) from the nation (public) when comparing one state to the nation at a time.
[1] Department of Defense Education Activity (overseas and domestic schools).
NOTE: States whose Black or White student population size was insufficient for comparison are omitted. Reporting standards not met for District of Columbia, Hawaii, Idaho, Maine, Montana, New Hampshire, North Dakota, South Dakota, Utah, Vermont, and Wyoming.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessment.

Trends in state mathematics achievement gaps at grade 8, 1990–2007

The national Black-White mathematics gap was not significantly narrower in 2007 than in 1990, despite higher average scores for both Black and White students in 2007 (figure 12, National results). The gap was narrower in 2007 than in 2005.

In 26 states, mathematics scores of both Black and White eighth-graders were higher in 2007 than in 1990. The 2007 gap was narrower in Arkansas, Colorado, Oklahoma, and Texas, as increases in Black students' scores were greater than those of their White peers.

Between 2005 and 2007, gaps narrowed in Arkansas and Florida as scores for Black eighth-graders increased while those of their White peers showed no change. In Colorado, scores for both groups increased, but a greater increase in Black students' scores caused the gap to narrow.

Narrowing of the Gap

In the following four states, the mathematics gap narrowed between **1990 and 2007** as gains of Black students outpaced the gains of White students.

Arkansas	**Oklahoma**
Colorado	**Texas**

In **Colorado**, the gap narrowed between **2005 and 2007** as Black students' scores showed greater increases than those of their White peers.

In **Arkansas** and **Florida**, the gap narrowed between **2005 and 2007** as Black students' scores increased while those of White students did not change significantly.
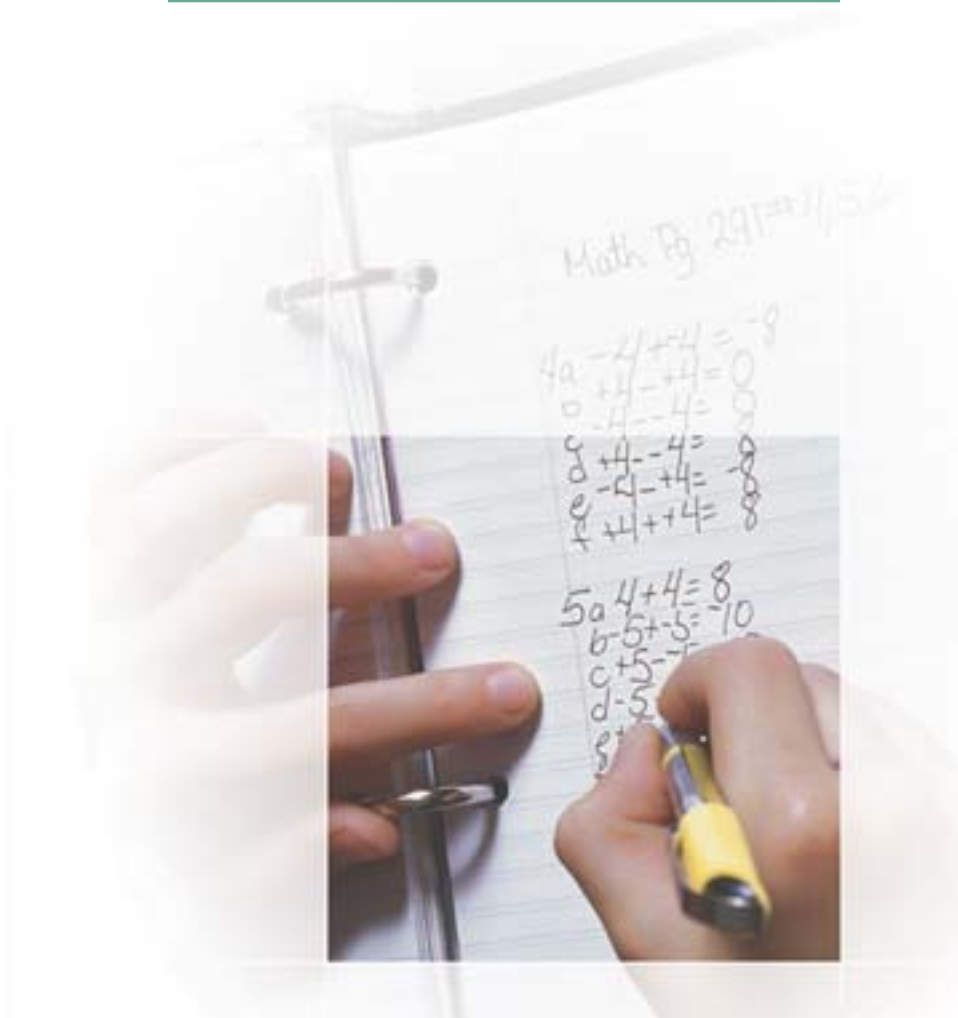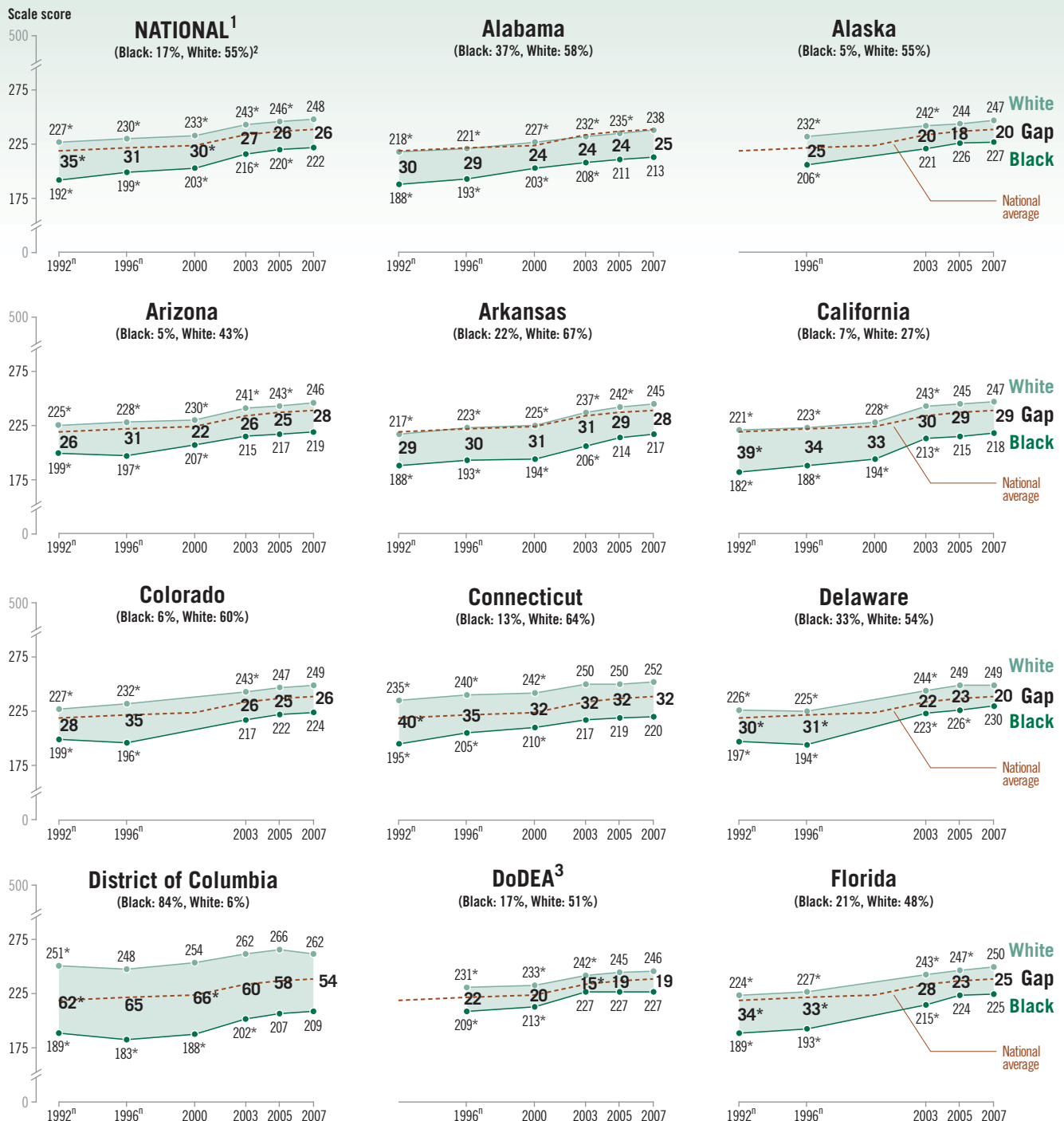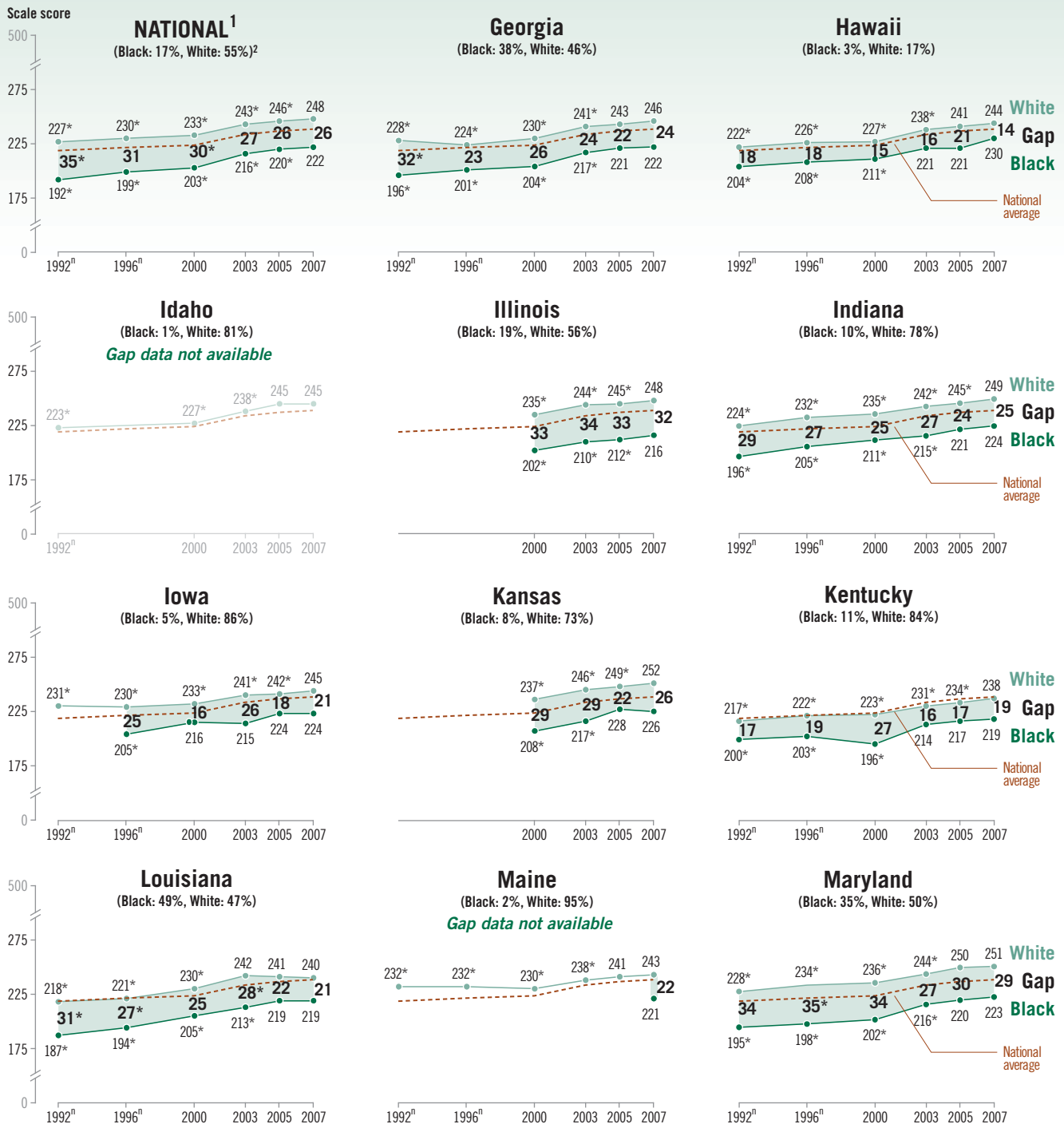
Figure 12. Gaps in average mathematics scores between Black and White public school students at grade 8, by state: Various years, 1990–2007

Scale score

NATIONAL[1]
(Black: 17%, White: 58%)[2]

White: 269* 276* 280* 283* 287* 288* 290
Gap: 33 40* 39* 40* 35* 33* 31
Black: 236* 236* 241* 243* 252* 254* 259

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Alabama
(Black: 35%, White: 60%)

262* 264* 270* 275 274 276 278
30 34 38 35 34 36 32
232* 230* 232* 240* 240* 240* 246

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Alaska
(Black: 4%, White: 56%)

285* 290* 288* 294 White
27 22 23 Gap
263 266 271 Black

National average

1996[n] 2003 2005 2007

Arizona
(Black: 5%, White: 47%)

270* 274* 277* 281* 284* 288 289 White
26 22 22 37 28 27 23 Gap
245* 253* 256 244* 256 261 266 Black

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Arkansas
(Black: 22%, White: 69%)

264* 264* 269* 268* 275* 281 282 White
34* 35* 35 41* 36* 38* 28 Gap
231* 229* 235* 227* 239* 243* 254 Black

National average

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

California
(Black: 7%, White: 31%)

270* 275* 277* 277* 283* 284* 287 White
38 42 33 42 37 35 35 Gap
231* 233* 244 235* 246 248 253 Black

National average

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Colorado
(Black: 7%, White: 65%)

273* 278* 282* 292* 292* 296 White
36* 36* 26 37* 36* 24 Gap
238* 242* 255 255* 256* 272 Black

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Connecticut
(Black: 13%, White: 69%)

277* 283* 287* 291 293 293 293 White
37 41 43 45* 38 44 38 Gap
240* 242* 244* 247* 255 249 255 Black

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Delaware
(Black: 31%, White: 56%)

268* 272* 275* 287* 291* 294 White
27 30 31 26 27 29 Gap
241* 241* 244* 260* 264 265 Black

National average

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

District of Columbia
(Black: 88%, White: 3%)

317
300
69 76
229* 232* 230* 231* 240* 241* 245

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

DoDEA[3]
(Black: 18%, White: 48%)

282* 286* 293 292 291 White
28* 27* 23 25 19 Gap
254* 259* 270 267 272 Black

1996[n] 2000 2003 2005 2007

Florida
(Black: 23%, White: 48%)

265* 272* 277* 286 286 289 White
34 36* 42* 37* 35* 29 Gap
231* 236* 235* 249* 251* 259 Black

National average

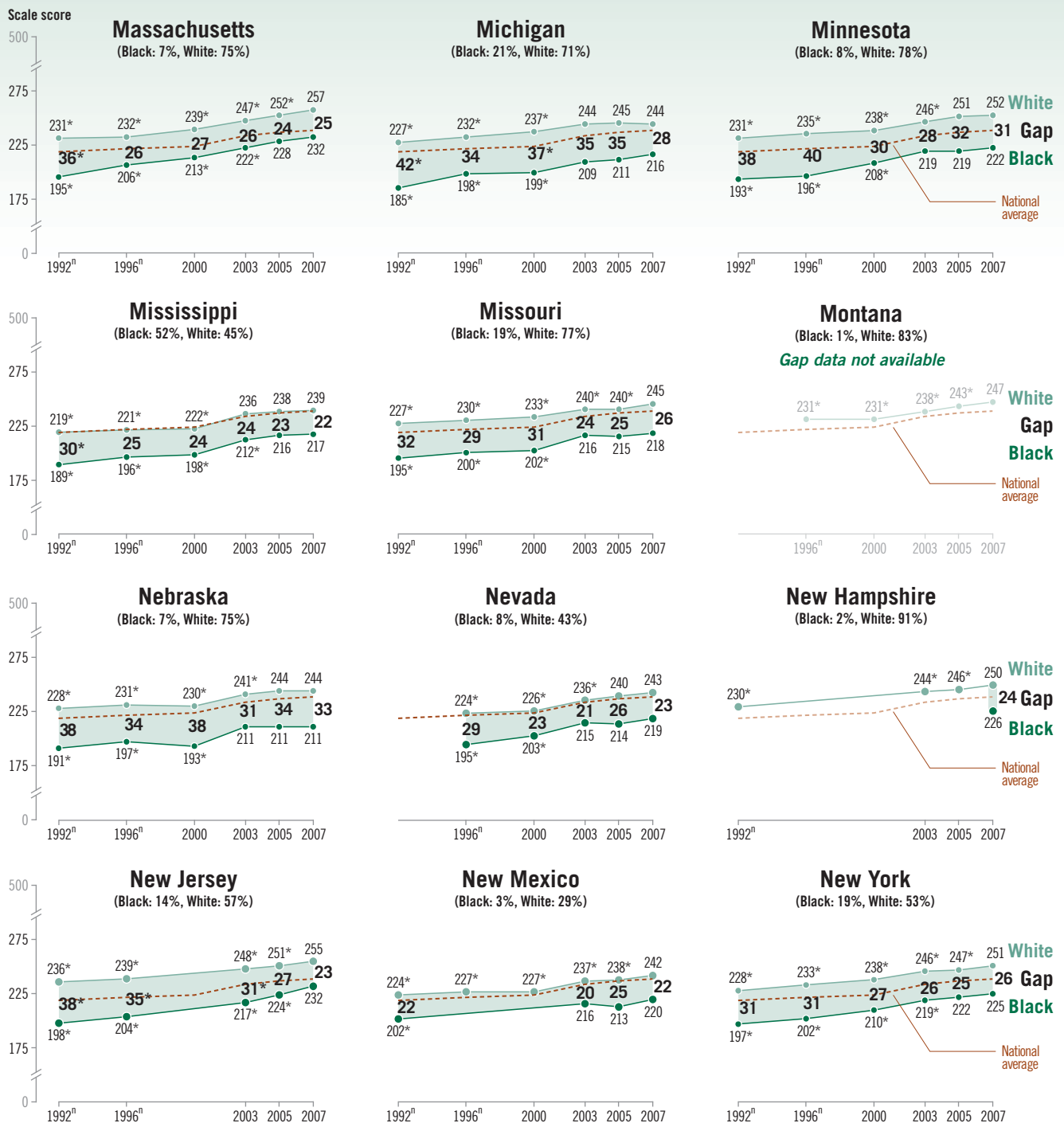1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

23

See notes at end of figure.

Figure 12. Gaps in average mathematics scores between Black and White public school students at grade 8, by state: Various years, 1990–2007—Continued
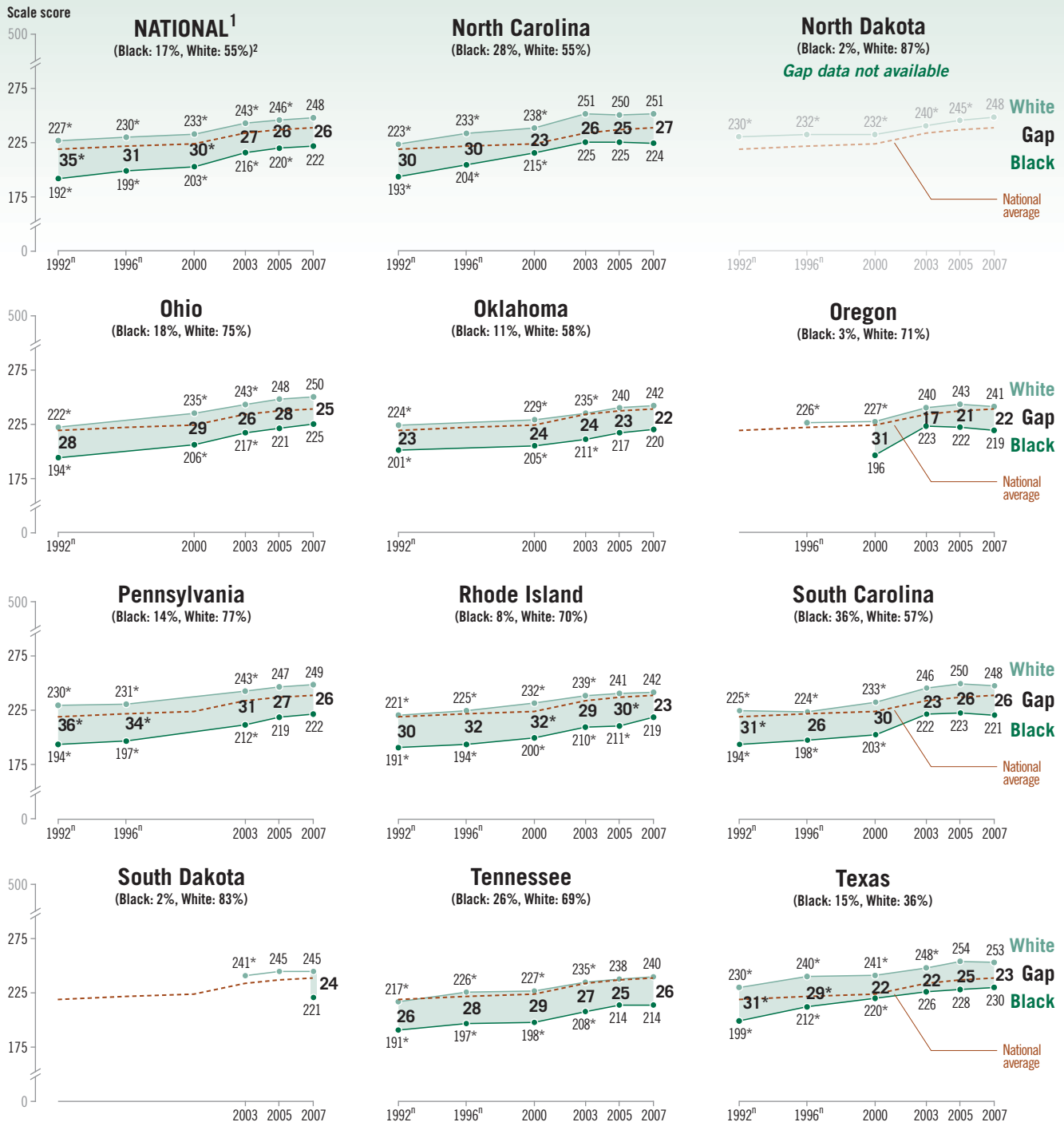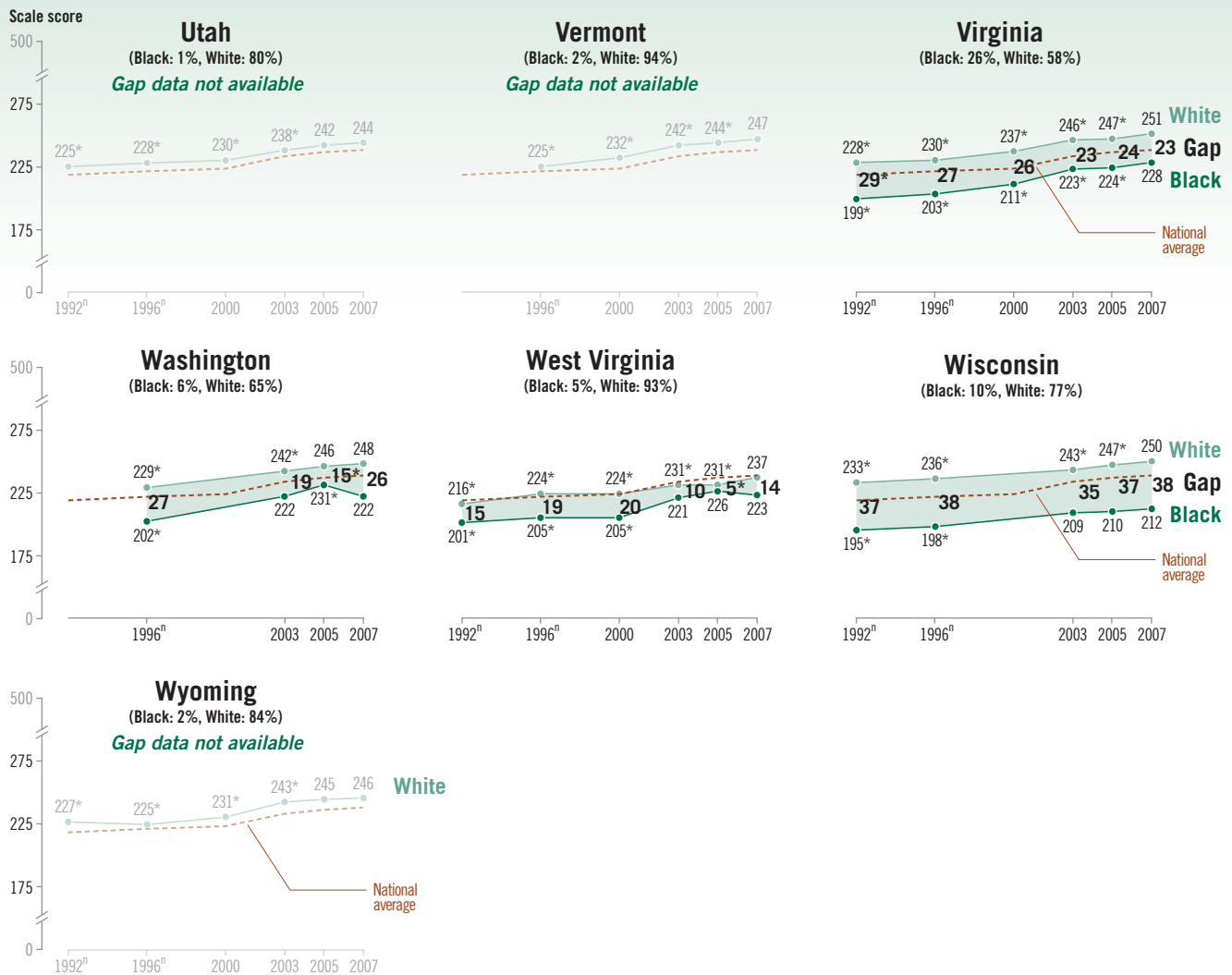

See notes at end of figure.

Figure 12. Gaps in average mathematics scores between Black and White public school students at grade 8, by state: Various years, 1990–2007—Continued

Scale score

Massachusetts
(Black: 8%, White: 75%)

White: 277* 283* 284* 292* 297* 305
Gap: 34 33 26* 33* 34 40
Black: 243* 250* 258 260 263 264
1992[n] 1996[n] 2000 2003 2005 2007

Michigan
(Black: 18%, White: 75%)

White: 270* 276* 284 285 286 285 285
Gap: 39 44 39 45 41 38 41
Black: 231* 233* 245 239 245 247 244
1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Minnesota
(Black: 7%, White: 81%)

White: 277* 284* 287* 290* 295 296 297
Gap: 41 39 44 46 37
Black: 236* 248 251 251 260
1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Mississippi
(Black: 51%, White: 47%)

White: 262* 265* 268* 275* 279 279
Gap: 33* 31 30 29 32 28
Black: 230* 234* 237* 246* 247* 251
1992[n] 1996[n] 2000 2003 2005 2007

Missouri
(Black: 19%, White: 75%)

White: 275* 278* 277* 284* 284* 288
Gap: 34 34 39 34 36 34
Black: 242* 244 238* 250 247 253
1992[n] 1996[n] 2000 2003 2005 2007

Montana
(Black: 1%, White: 85%)
Gap data not available

White: 282* 286* 288* 289* 290 291
1990[n] 1996[n] 2000 2003 2005 2007

Nebraska
(Black: 7%, White: 80%)

White: 279* 281* 285* 285* 287* 289 291
Gap: 45 44 32* 37* 41* 46 51
Black: 234 237 254* 247 247 243 240
1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Nevada
(Black: 10%, White: 47%)

White: 273* 278 280 282
Gap: 29 30 33 26
Black: 244 248 247 255
2000 2003 2005 2007

New Hampshire
(Black: 2%, White: 94%)
Gap data not available

White: 273* 278* 287 286* 289
1990[n] 1992[n] 2003 2005 2007

New Jersey
(Black: 17%, White: 57%)

White: 279* 283* 292* 295* 298
Gap: 38 41 39 35 35
Black: 241* 242* 253* 260 264
1990[n] 1992[n] 2003 2005 2007

New Mexico
(Black: 3%, White: 32%)

White: 271* 272* 277* 274* 282 279* 285
Gap: 28 22 21
Black: 254 257 264
1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

New York
(Black: 19%, White: 55%)

White: 273* 280* 283* 284* 293 290 290
Gap: 39 46* 40 33 37 32 32
Black: 234* 233* 243* 251 255 259 258
1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

White
Gap
Black
National average

See notes at end of figure.

Figure 12. Gaps in average mathematics scores between Black and White public school students at grade 8, by state: Various years, 1990–2007—Continued

Scale score

NATIONAL[1]
(Black: 17%, White: 58%)[2]

White: 269*, 276*, 280*, 283*, 287*, 288*, 290
Gap: 33, 40*, 39*, 40*, 35*, 33*, 31
Black: 236*, 236*, 241*, 243*, 252*, 254*, 259

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

North Carolina
(Black: 30%, White: 56%)

White: 261*, 266*, 277*, 287*, 294, 292*, 295
Gap: 30, 28, 30, 35*, 34, 28, 29
Black: 231*, 238*, 247*, 252*, 260*, 263, 266

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

North Dakota
(Black: 1%, White: 89%)
Gap data not available

White: 284*, 284*, 286*, 285*, 290*, 290*, 295

National average

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Ohio
(Black: 18%, White: 76%)

White: 268*, 274*, 285*, 287*, 289, 291
Gap: 35, 41, 34, 30, 34, 33
Black: 233*, 234*, 251, 257, 255, 258

1990[n] 1992[n] 2000 2003 2005 2007

Oklahoma
(Black: 9%, White: 59%)

White: 268*, 272*, 274*, 278, 278, 280
Gap: 32*, 34*, 29, 29, 29, 22
Black: 236*, 238*, 245*, 249, 249*, 258

1990[n] 1992[n] 2000 2003 2005 2007

Oregon
(Black: 3%, White: 73%)

White: 273*, 278*, 284*, 284*, 287, 289
Gap: 20, 29, 16
Black: 265, 258, 272

National average

1990[n] 1996[n] 2000 2003 2005 2007

Pennsylvania
(Black: 15%, White: 76%)

White: 272*, 276*, 285*, 287*, 293
Gap: 36, 39, 38, 37, 36
Black: 236*, 238*, 247*, 250, 257

1990[n] 1992[n] 2003 2005 2007

Rhode Island
(Black: 9%, White: 70%)

White: 265*, 271*, 275*, 275*, 280*, 281*, 284
Gap: 37, 30, 38, 35, 36, 32, 34
Black: 228*, 240, 237*, 240, 244, 249, 250

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

South Carolina
(Black: 38%, White: 56%)

White: 273*, 273*, 277*, 291, 294, 293
Gap: 33*, 29, 30, 33*, 32, 27
Black: 241*, 244*, 247*, 258*, 263, 265

National average

1992[n] 1996[n] 2000 2003 2005 2007

South Dakota
(Black: 1%, White: 86%)
Gap data not available

White: 288*, 291, 292

2003 2005 2007

Tennessee
(Black: 28%, White: 67%)

White: 266*, 270*, 269*, 277*, 278*, 282
Gap: 32, 35, 34, 35*, 32, 28
Black: 234*, 234*, 235*, 242*, 246*, 254

1992[n] 1996[n] 2000 2003 2005 2007

Texas
(Black: 15%, White: 38%)

White: 272*, 278*, 284*, 286*, 290*, 295*, 300
Gap: 38*, 35, 35, 37, 30, 31, 29
Black: 234*, 243*, 249*, 250*, 260*, 264*, 271

National average

1990[n] 1992[n] 1996[n] 2000 2003 2005 2007

Legend: White / Gap / Black / National average

See notes at end of figure.

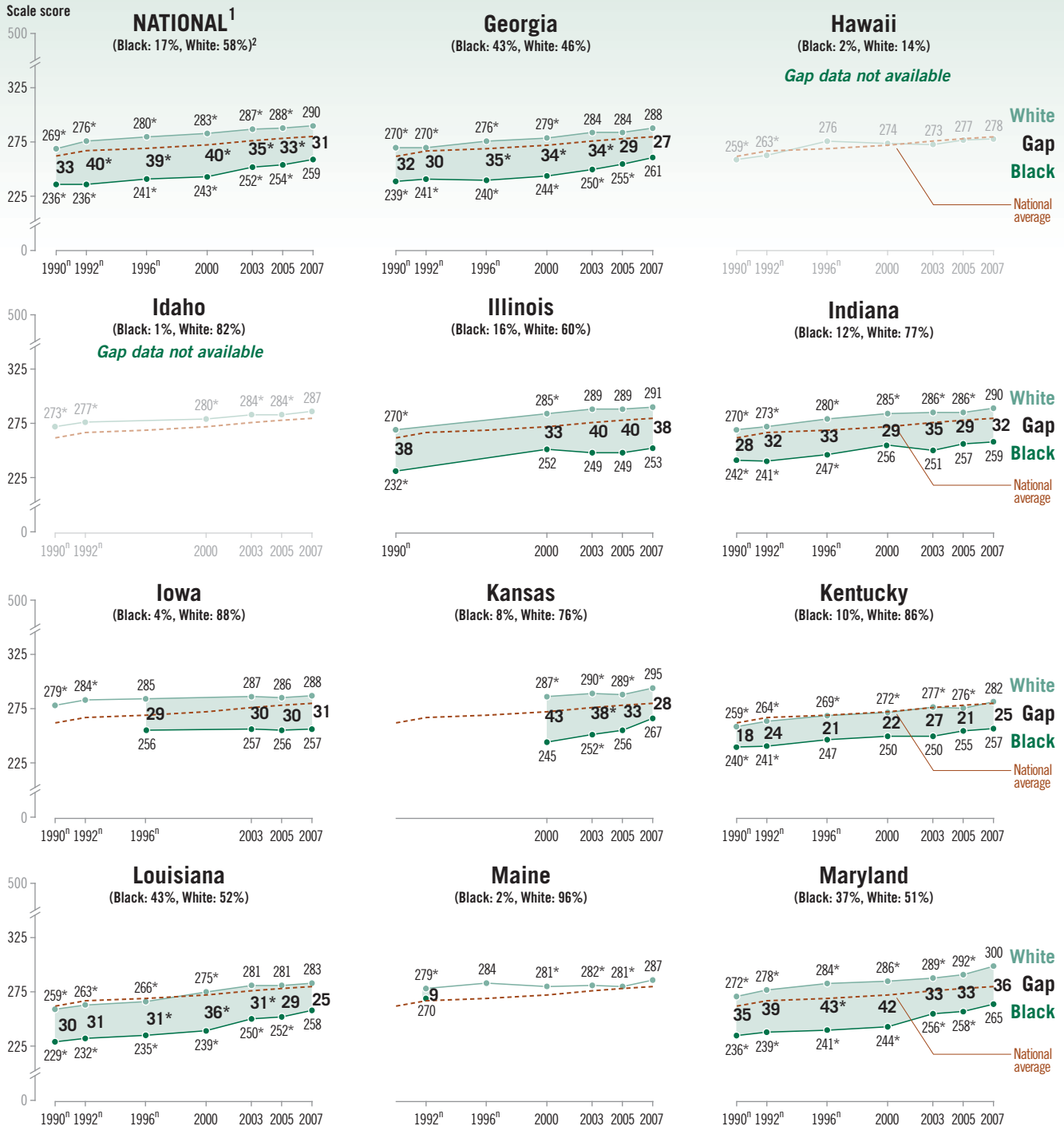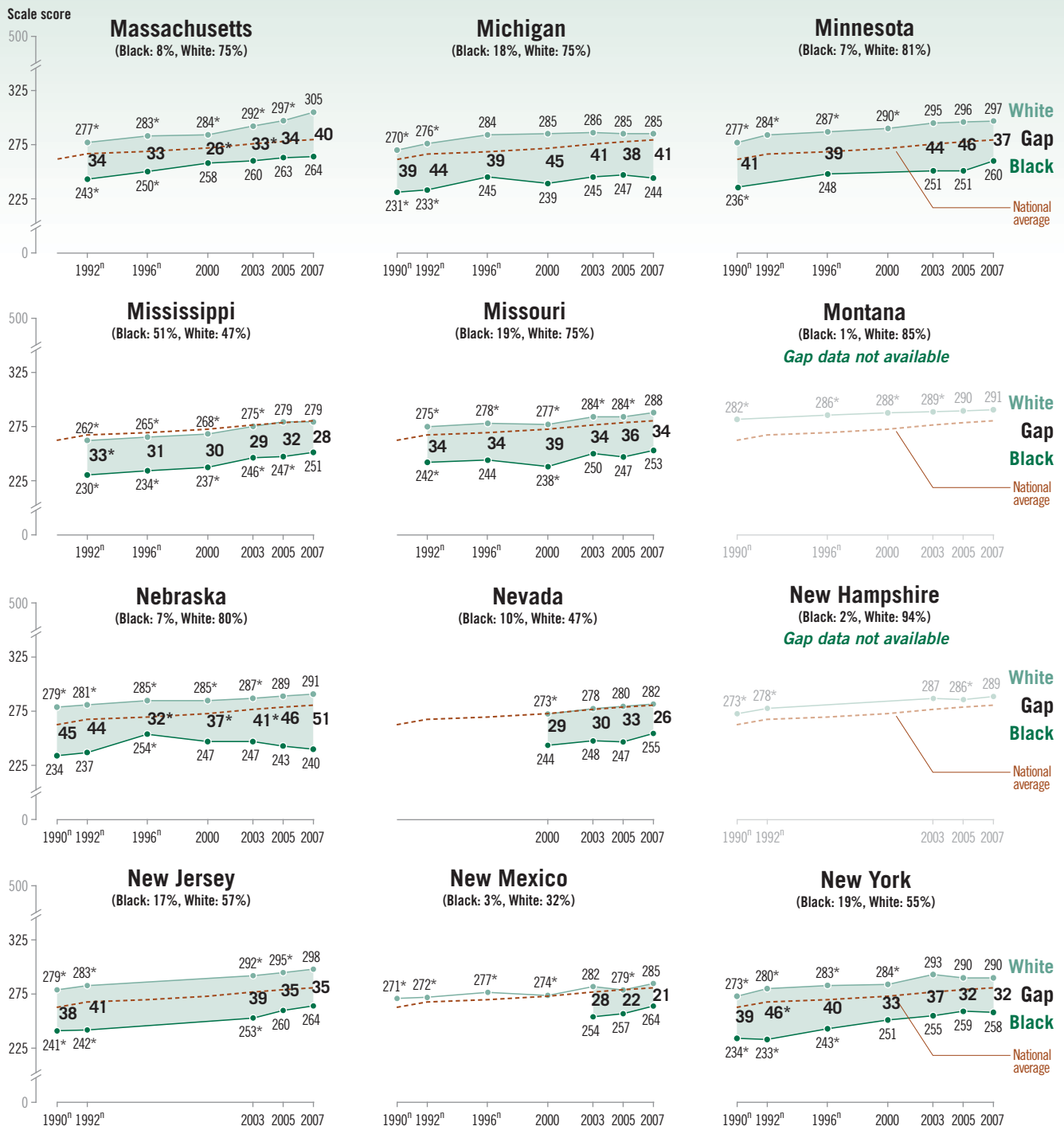**Figure 12. Gaps in average mathematics scores between Black and White public school students at grade 8, by state: Various years, 1990–2007—Continued**
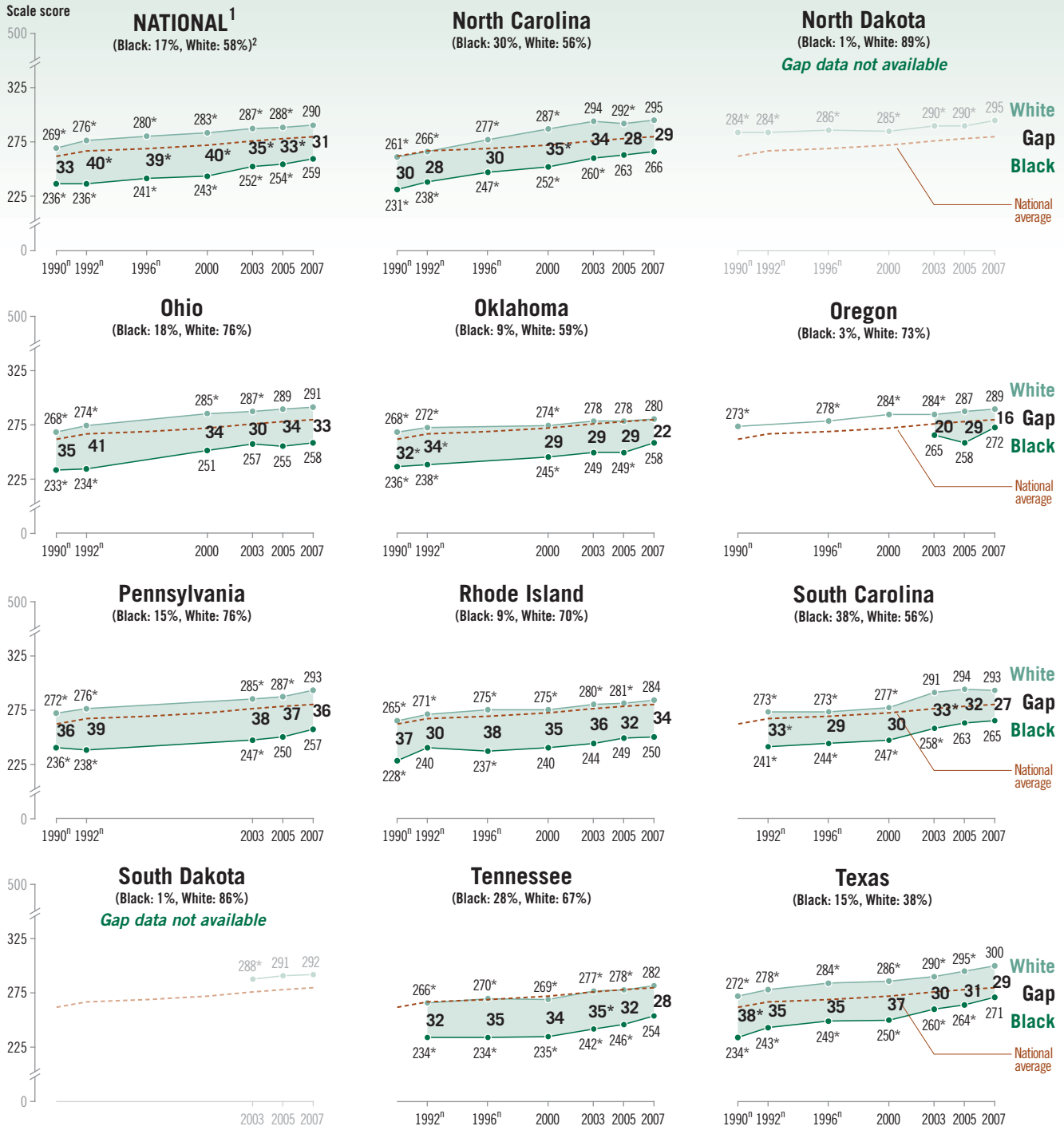

Scale score

Utah
(Black: 1%, White: 82%)
Gap data not available

Vermont
(Black: 1%, White: 95%)
Gap data not available

Virginia
(Black: 26%, White: 61%)

Washington
(Black: 5%, White: 69%)

West Virginia
(Black: 4%, White: 94%)

Wisconsin
(Black: 10%, White: 80%)

Wyoming
(Black: 1%, White: 86%)
Gap data not available

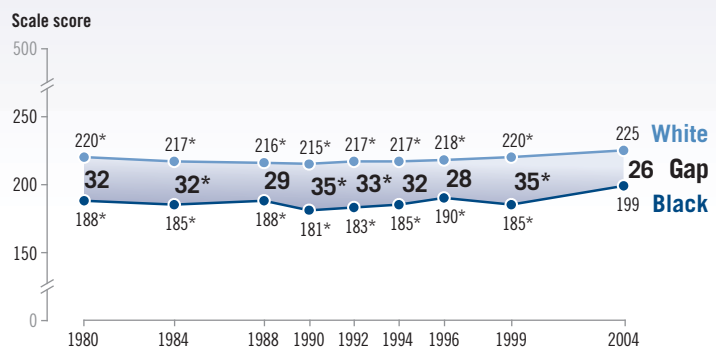# Long-Term Trend Results for Black and White 9- and 13-Year-Olds

Trends in reading scores and achievement gaps, 1980–2004

Reading scores for both Black and White 9-year-old students were higher in 2004 than on any previous long-term trend assessment (figure 13). The score gap in 2004 did not differ significantly from the gap in 1980, but was narrower than the gap in 1999, due to a greater increase in Black students' scores as compared to White students.
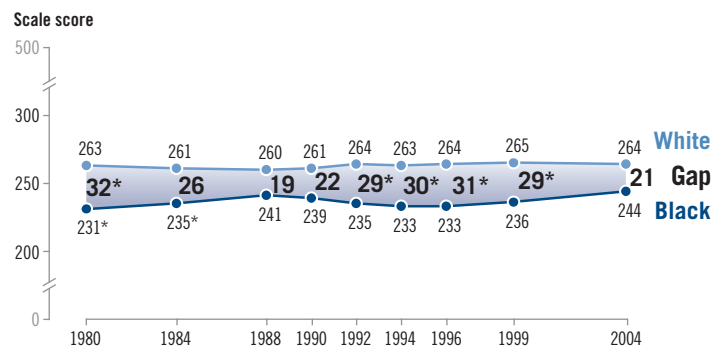
At age 13, reading scores for White students were not significantly different in 2004 than in 1980 (figure 14). For Black students, scores were higher in 2004 than in 1980, resulting in a narrowing of the gap. Scores did not change significantly for either Black or White students from 1999 to 2004, but the gap narrowed for that time period as well. A statistically significant change can occur over time in the gap between two scores even though the scores themselves do not change significantly because changes in gaps are calculated separately from changes in scores.

Figure 13. Trends in average reading scale scores and score gaps for White students and Black students at age 9: Various years, 1980–2004

Scale score

White	220*	217*	216*	215*	217*	217*	218*	220*	225
Gap	32	32*	29	35*	33*	32	28	35*	26
Black	188*	185*	188*	181*	183*	185*	190*	185*	199

1980 1984 1988 1990 1992 1994 1996 1999 2004

* Significantly different (p<.05) from 2004.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1980–2004 Long-Term Trend Reading Assessments.

Figure 14. Trends in average reading scale scores and score gaps for White students and Black students at age 13: Various years, 1980–2004

Scale score

White	263	261	260	261	264	263	264	265	264
Gap	32*	26	19	22	29*	30*	31*	29*	21
Black	231*	235*	241	239	235	233	233	236	244

1980 1984 1988 1990 1992 1994 1996 1999 2004

* Significantly different (p<.05) from 2004.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1980–2004 Long-Term Trend Reading Assessments.

Main NAEP National Results for Black and White Fourth- and Eighth-Graders

Trends in reading scores and achievement gaps, 1992–2007

In main NAEP, the reading gap for Black and White fourth-graders narrowed in 2007 in comparison to both 1992 and 2005 (figure 15). Although scores for both Black and White students were higher in 2007 than in either comparison year, a greater increase in scores for Black students caused the gap to narrow. The 27-point gap in 2007 was narrower than in any previous assessment year except 1998.

Eighth-grade reading scores for both Black and White students were higher in 2007 than in either 1992 or 2005, but the gap in 2007 was not significantly different from either prior year (figure 16).

Figure 15. Reading achievement score gaps between Black and White public school students at grade 4: Various years, 1992–2007
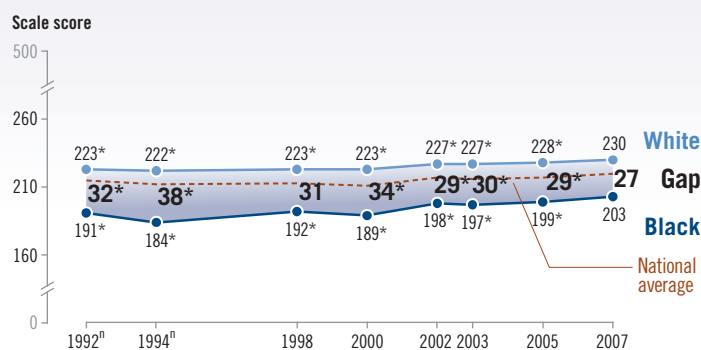
[n] Accommodations were not permitted for this assessment.
* Significantly different (p<.05) from 2007.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1992–2007 Reading Assessments.

Figure 16. Reading achievement score gaps between Black and White public school students at grade 8: Various years, 1992–2007

[n] Accommodations were not permitted for this assessment.
* Significantly different (p<.05) from 2007.
NOTE: Data were not collected at grade 8 in 2000.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1992–2007 Reading Assessments.

Reading scores and achievement gaps by gender, 1992–2007

Average reading scores were higher in 2007 than in 1992 for Black and for White fourth-graders, regardless of gender (figure 17). Among males, the gap narrowed as the scores of Black males increased more than those of their White peers. Among fourth-grade females, the Black-White gap did not change significantly.

Fourth-grade reading scores were higher in 2007 than in 2005 for both Black and White males and females, and the achievement gaps narrowed for both groups during this period, as the scores of Black fourth-graders increased more than those of their White peers.

Average reading scores for eighth-graders were higher in 2007 than in 1992 for Black and for White students, regardless of gender (figure 18). However, the 2007 gaps in eighth-grade reading achievement showed no significant differences from the 1992 gaps for either males or females.

From 2005 to 2007, average reading scores for eighth-graders increased for both Black and White males. Scores increased for Black females but not for White females. However, the Black-White gap did not change significantly for either gender during this period.

Figure 17. Gaps in average reading scores between Black and White public school students at grade 4, by gender: Various years, 1992–2007
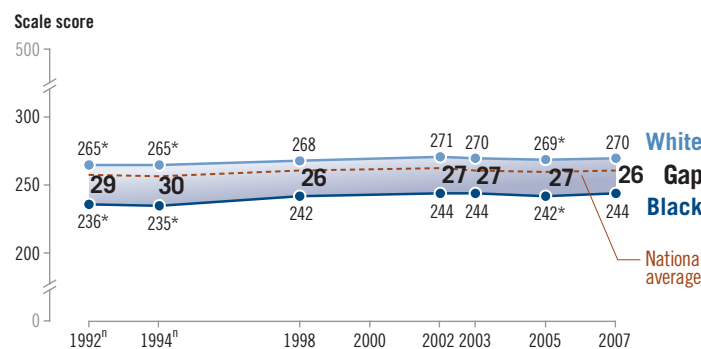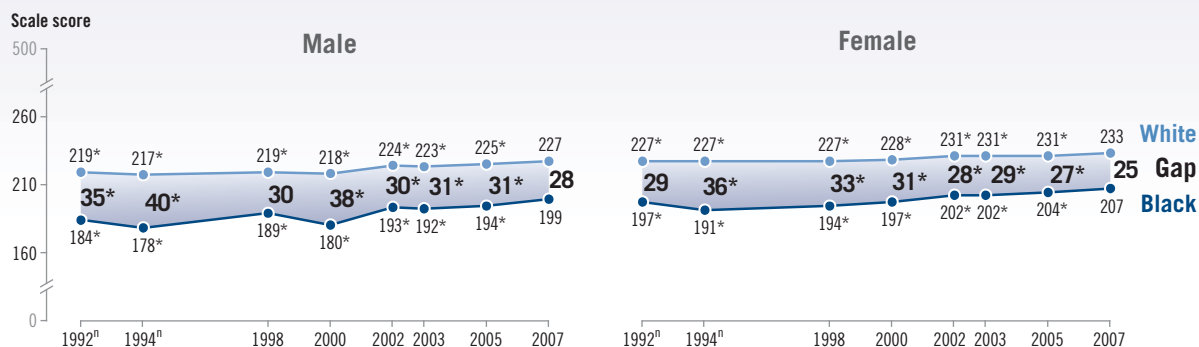

[n] Accommodations were not permitted for this assessment.
* Significantly different (p<.05) from 2007.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1992–2007 Reading Assessments.

Figure 18. Gaps in average reading scores between Black and White public school students at grade 8, by gender: Various years, 1992–2007

[n] Accommodations were not permitted for this assessment.
* Significantly different (p<.05) from 2007.
NOTE: Data were not collected at grade 8 in 2000.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1992–2007 Reading Assessments.

Reading scores and achievement gaps by family income, 2003–2007

NAEP uses student eligibility for free or reduced-price school lunch as an indicator of family income. At grade 4, reading scores were higher in 2007 than in 2003 for both Black and White public school students, regardless of school-lunch eligibility (figure 19). The gap in 2007 for not-eligible students was narrower than in 2003, while the gap for students eligible for free lunch was narrower than in either previous assessment.

At grade 8, scores were higher for Black and White not-eligible students only, comparing 2007 with 2005 (figure 20). There were no statistically significant changes in the sizes of the gaps.

Eligibility for free and reduced-price lunch

NAEP collects data on students' eligibility for the National School Lunch Program (NSLP)—sometimes referred to as the free and reduced-price school lunch program—as an indicator of family economic status. Eligibility for free and reduced-price lunches is based on students' family income in relation to the federally established poverty level.

Not eligible: Students who are not eligible for the program because their family's income is above 185 percent of the poverty level.

Eligible for reduced-price lunch: Students who are eligible for reduced-price lunch because their family's income is between 130 percent and 185 percent of the poverty level.

Eligible for free lunch: Students who are eligible for free lunch because their family's income is below 130 percent of the poverty level.

As a result of improvements in the quality of the data on students' eligibility for NSLP, the percentage of students for whom information was not available has decreased in comparison to the percentages reported prior to the 2003 assessment. Therefore, trend comparisons are only made back to 2003 in this report.

Table 2. **Percentage of public school students assessed in NAEP reading by eligibility for free or reduced-price school lunch, race/ethnicity and grade: 2003, 2005, and 2007**

	Not eligible		Eligible for reduced-price lunch		Eligible for free lunch	
	Black	White	Black	White	Black	White
Grade 4						
2007	26	73	7	6	66	21
2005	25	72	8	7	66	20
2003	24	72	9	8	65	18
Grade 8						
2007	32	76	7	5	59	18
2005	32	75	9	6	57	17
2003	32	76	9	6	56	14

NOTE: Detail may not sum to totals due to rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Reading Assessments.

Figure 19.   Gaps in average reading scores between Black and White public school students at grade 4, by eligibility for free or reduced-price school lunch: 2003, 2005, and 2007

Scale score

Not eligible

Eligible for reduced-price lunch

Eligible for free lunch

White
Gap
Black
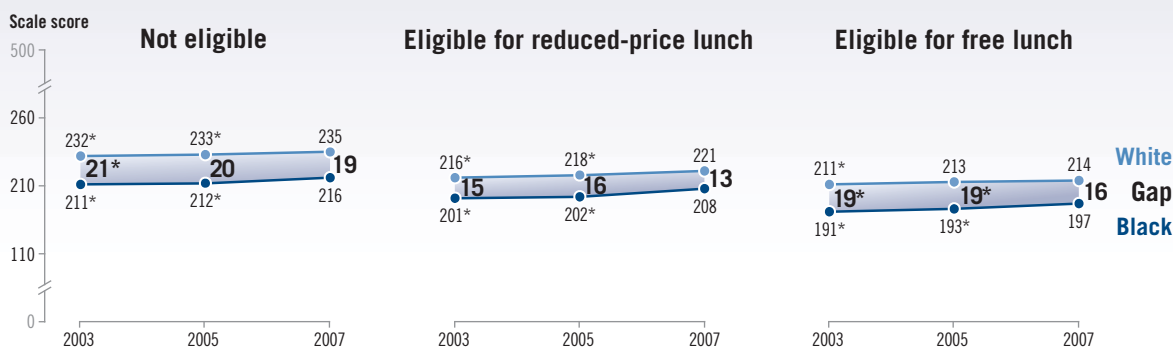
* Significantly different (p<.05) from 2007.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Reading Assessments.

Figure 20.   Gaps in average reading scores between Black and White public school students at grade 8, by eligibility for free or reduced-price school lunch: 2003, 2005, and 2007

Scale score

Not eligible

Eligible for reduced-price lunch

Eligible for free lunch

White
Gap
Black
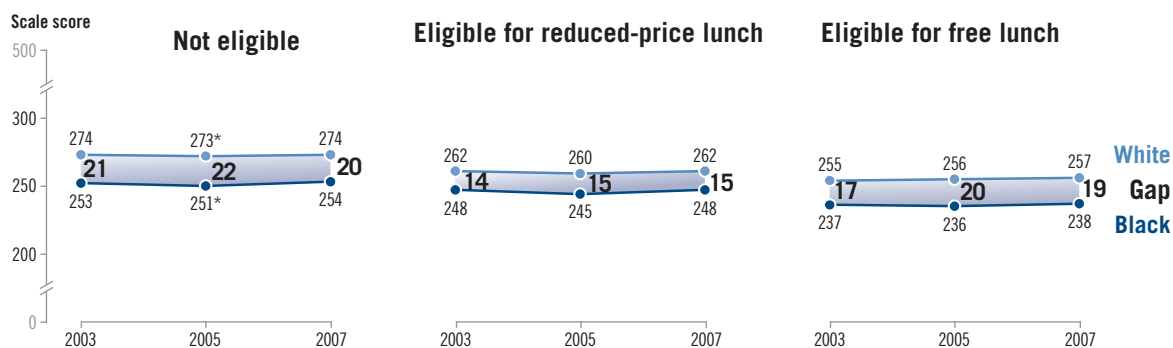
* Significantly different (p<.05) from 2007.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003, 2005, and 2007 Reading Assessments.

Main NAEP State Results for Black and White Fourth- and Eighth-Graders

The NAEP state reading assessments were administered to public school students in fourth grade in 1992, 1994, 1998, 2002, 2003, 2005, and 2007 and in eighth grade in 1998, 2002, 2003, 2005, and 2007. Before 2003, states were not required to participate in NAEP in order to qualify for Title I education funds. Typically, 40 or more states participated in each assessment prior to 2003. In 2003, 2005, and 2007, all 50 states, the District of Columbia, and the DoDEA schools participated.

State results are presented in two ways. Comparisons of fourth-grade reading gaps in 2007 between each state and the nation are presented in figure 21.

Comparisons of the reading gaps within a state over time are presented in a series of small graphs in figure 22. At the top left of each two-page spread, the reading scores and gaps for the nation are presented for reference. Each state figure, as well as the national figure, also contains a dotted red line representing the national average for public school students. The data for the national averages are located in appendix B in table B-4.
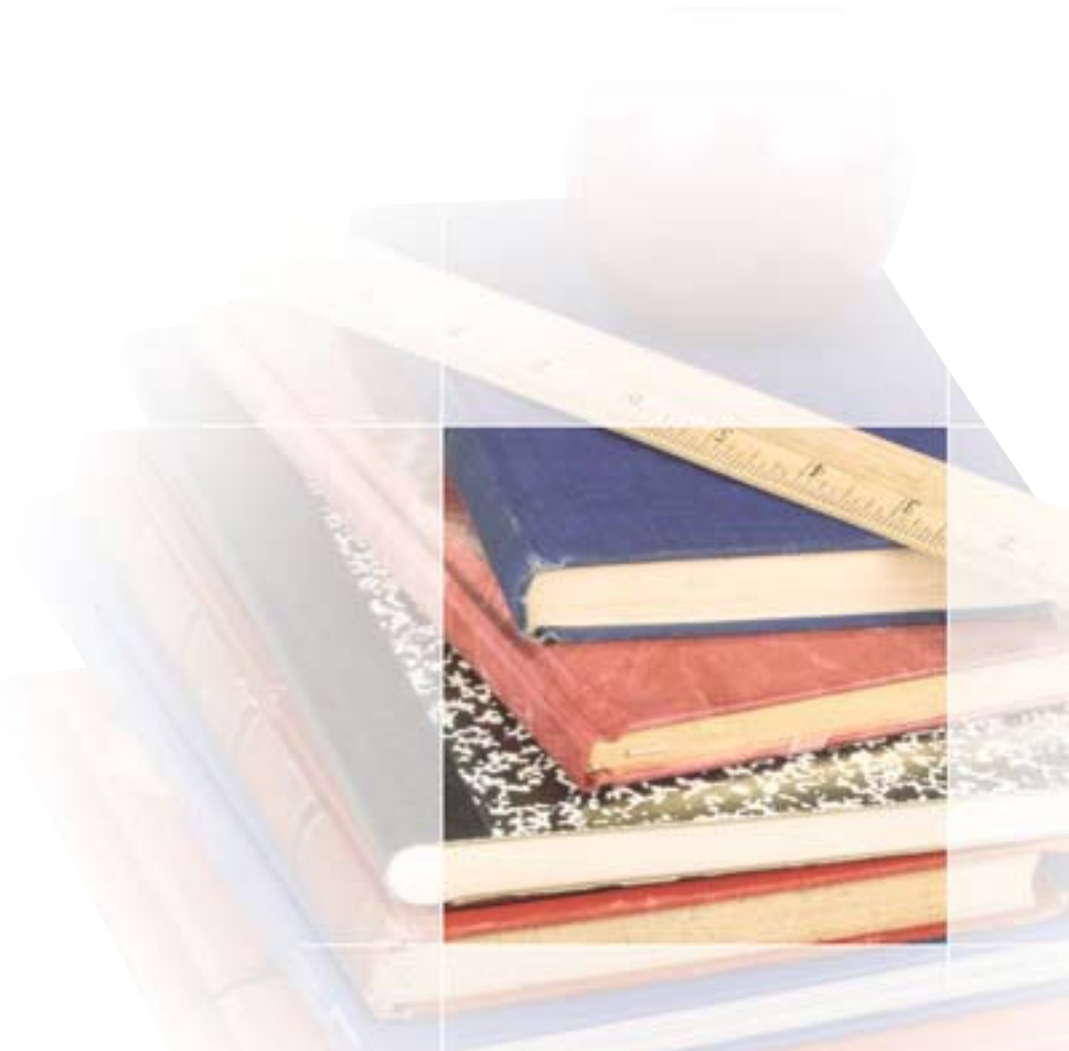
State and national reading achievement gaps at grade 4, 2007

Nine states had a Black-White gap that was smaller than the nation's 27-point gap in 2007 (Arizona, Delaware, DoDEA, Hawaii, Kentucky, New Hampshire, Oklahoma, Virginia, and West Virginia) and eight had a gap that was larger (Arkansas, Connecticut, District of Columbia, Minnesota, Nebraska, Pennsylvania, Tennessee, and Wisconsin). In 27 states, the gap was not different from the national gap. Gaps that are different from the national gap are indicated with an asterisk (figure 21).

The Black-White grade 4 reading gap in 2007 was significant in all 44 states for which data could be reported. The gaps ranged from 13 points in West Virginia to 67 points in the District of Columbia.

Figure 21. The Black-White achievement score gap in reading for public school students at grade 4, by state or jurisdiction: 2007
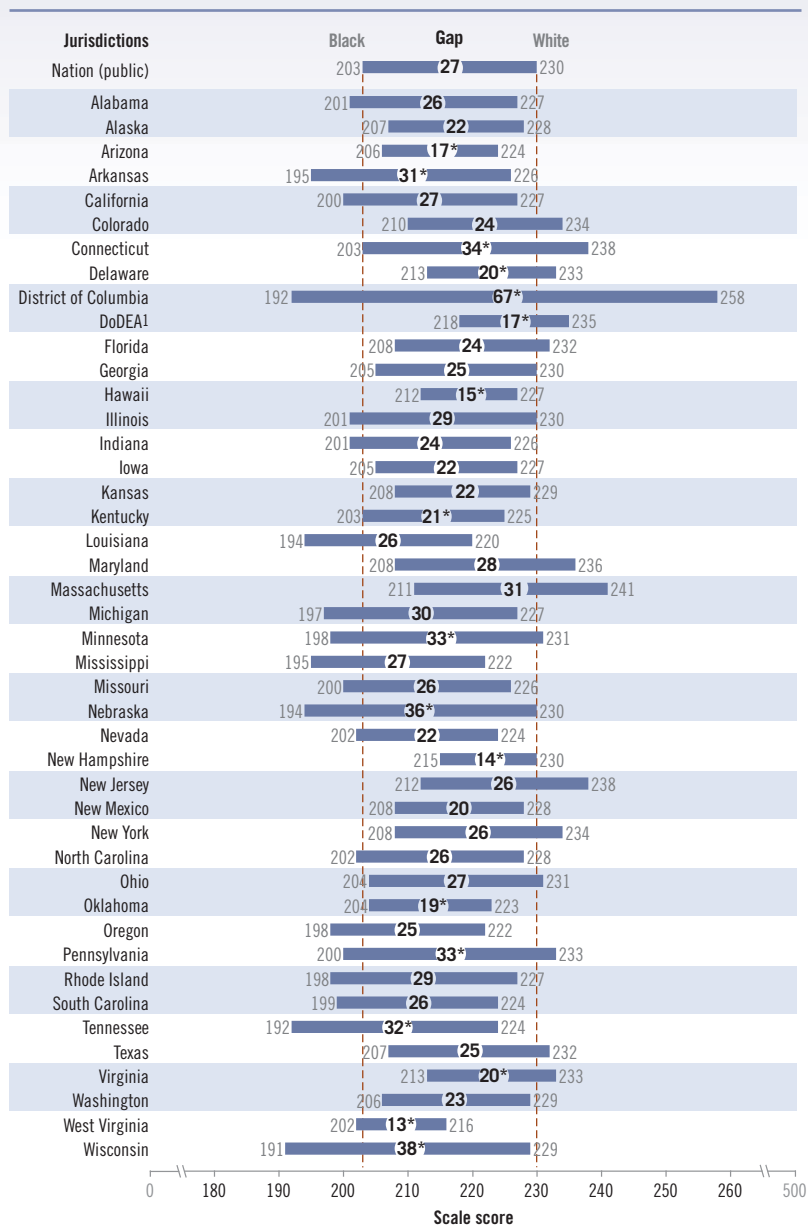
* Significantly different (p<.05) from the nation (public) when comparing one state to the nation at a time.
[1] Department of Defense Education Activity (overseas and domestic schools).
NOTE: States whose Black student population size was insufficient for comparison are omitted. Reporting standards not met for Idaho, Maine, Montana, North Dakota, South Dakota, Utah, Vermont, and Wyoming.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessment.

Trends in state reading achievement gaps at grade 4, 1992–2007

The Black-White reading gap among the nation's public school fourth-graders was narrower in 2007 than in 1992 as average scores for Black students demonstrated a larger increase than average scores for White students (figure 22, National results).

In 13 states, both Black and White fourth-graders achieved higher average scores in reading during this period. In three states—Delaware, Florida, and New Jersey—the gap was narrower in 2007 than in 1992 as Black students' scores increased more than those of White students.

In addition, gaps narrowed from 2005 to 2007 in Alabama, Arizona, and Virginia.

Narrowing of the Gap

In the following three states, the reading gap was narrower in **2007 than in 1992**, as Black students' scores increased more than those of their White peers.

Delaware   **New Jersey**
Florida

In **Alabama**, the reading gap narrowed between **2005 and 2007** as Black students' scores increased more than White students' scores.

In **Arizona and Virginia**, the reading gap narrowed between **2005 and 2007** as Black students' scores increased while those of White students did not change significantly.

Figure 22.   Gaps in average reading scores between Black and White public school students at grade 4, by state: Various years, 1992–2007

Scale score

NATIONAL[1]
(Black: 17%, White: 56%)[2]

Alabama
(Black: 37%, White: 58%)

Alaska
(Black: 4%, White: 54%)

Arizona
(Black: 5%, White: 44%)

Arkansas
(Black: 20%, White: 70%)

California
(Black: 7%, White: 28%)

Colorado
(Black: 5%, White: 62%)

Connecticut
(Black: 14%, White: 64%)

Delaware
(Black: 34%, White: 53%)

District of Columbia
(Black: 86%, White: 6%)

DoDEA[3]
(Black: 19%, White: 49%)

Florida
(Black: 21%, White: 47%)

See notes at end of figure.

37

Figure 22.   Gaps in average reading scores between Black and White public school students at grade 4, by state: Various years, 1992–2007—Continued

See notes at end of figure.

Figure 22. Gaps in average reading scores between Black and White public school students at grade 4, by state: Various years, 1992–2007—Continued

Massachusetts
(Black: 8%, White: 75%)

Michigan
(Black: 20%, White: 71%)

Minnesota
(Black: 8%, White: 78%)

Mississippi
(Black: 51%, White: 47%)

Missouri
(Black: 20%, White: 75%)

Montana
(Black: 1%, White: 83%)
Gap data not available

Nebraska
(Black: 8%, White: 76%)

Nevada
(Black: 9%, White: 44%)

New Hampshire
(Black: 2%, White: 92%)

New Jersey
(Black: 15%, White: 59%)

New Mexico
(Black: 3%, White: 32%)

New York
(Black: 19%, White: 53%)

See notes at end of figure.

39

Figure 22.   Gaps in average reading scores between Black and White public school students at grade 4, by state: Various years, 1992–2007—Continued

Scale score

NATIONAL[1]
(Black: 17%, White: 56%)[2]

North Carolina
(Black: 27%, White: 56%)

North Dakota
(Black: 2%, White: 88%)
Gap data not available

Ohio
(Black: 17%, White: 75%)

Oklahoma
(Black: 10%, White: 60%)

Oregon
(Black: 3%, White: 69%)

Pennsylvania
(Black: 15%, White: 76%)

Rhode Island
(Black: 9%, White: 68%)

South Carolina
(Black: 36%, White: 56%)

South Dakota
(Black: 2%, White: 84%)
Gap data not available

Tennessee
(Black: 25%, White: 70%)

Texas
(Black: 16%, White: 37%)
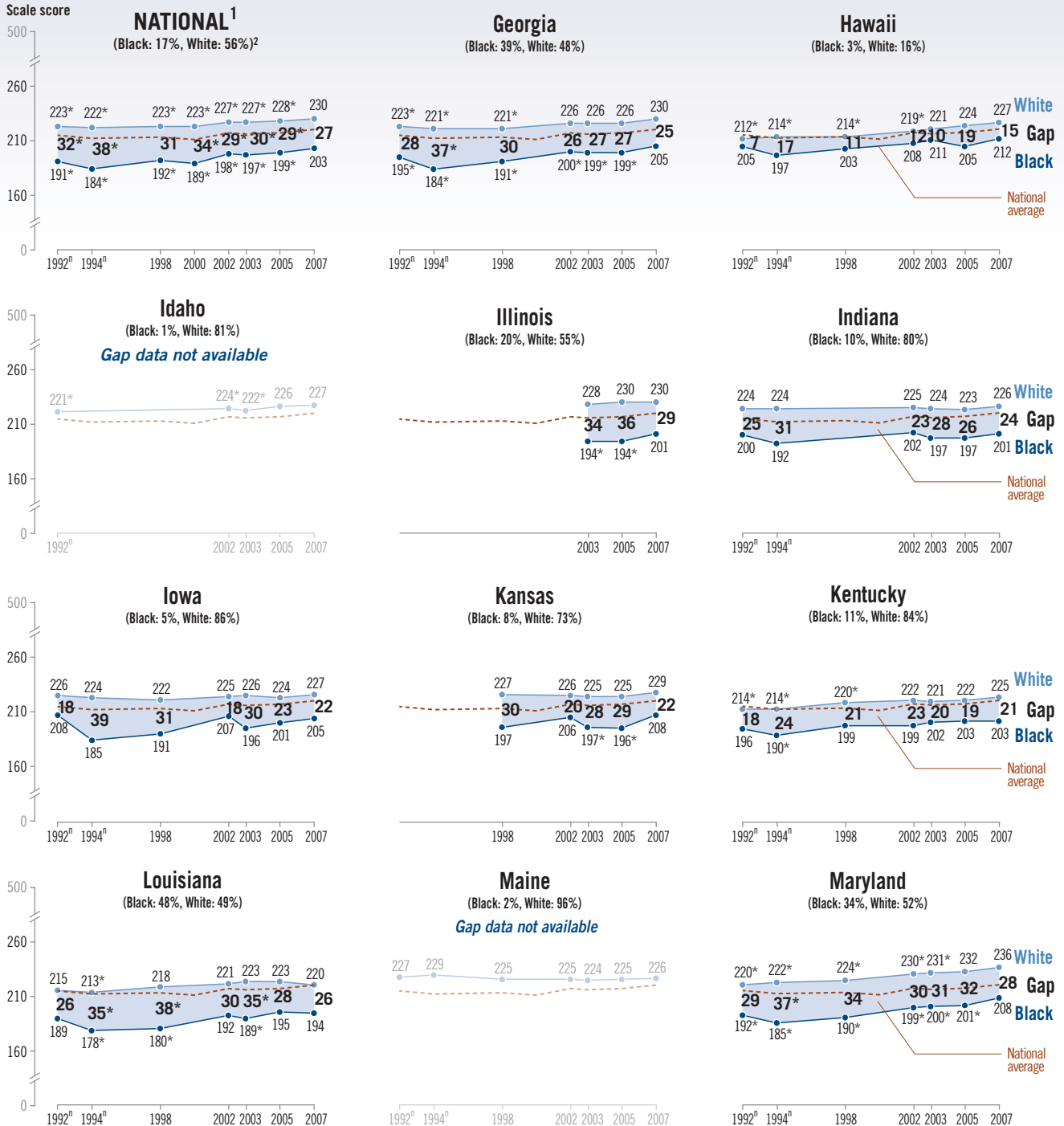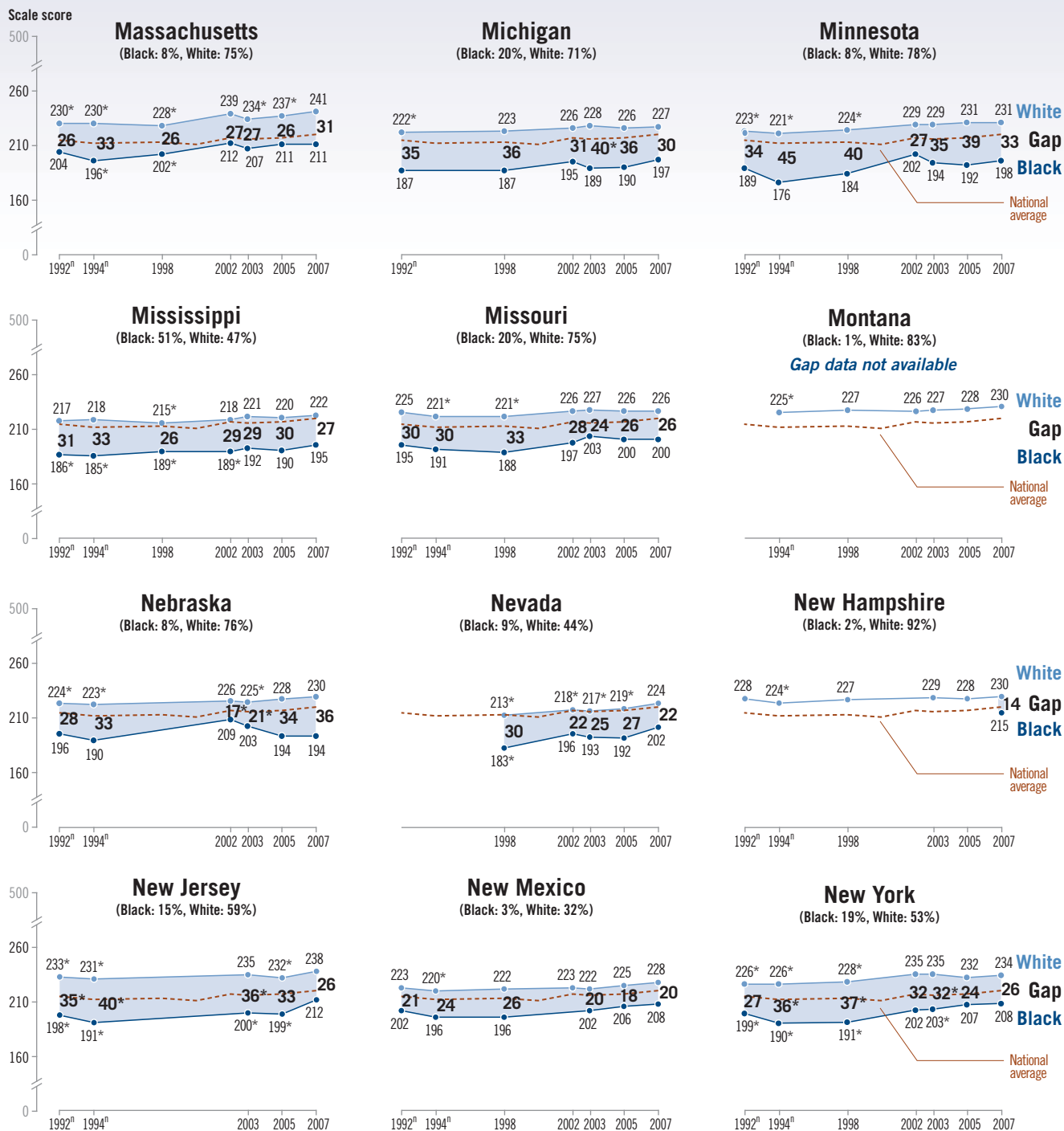
White
Gap
Black
National average

See notes at end of figure.

Figure 22. Gaps in average reading scores between Black and White public school students at grade 4, by state: Various years, 1992–2007—Continued
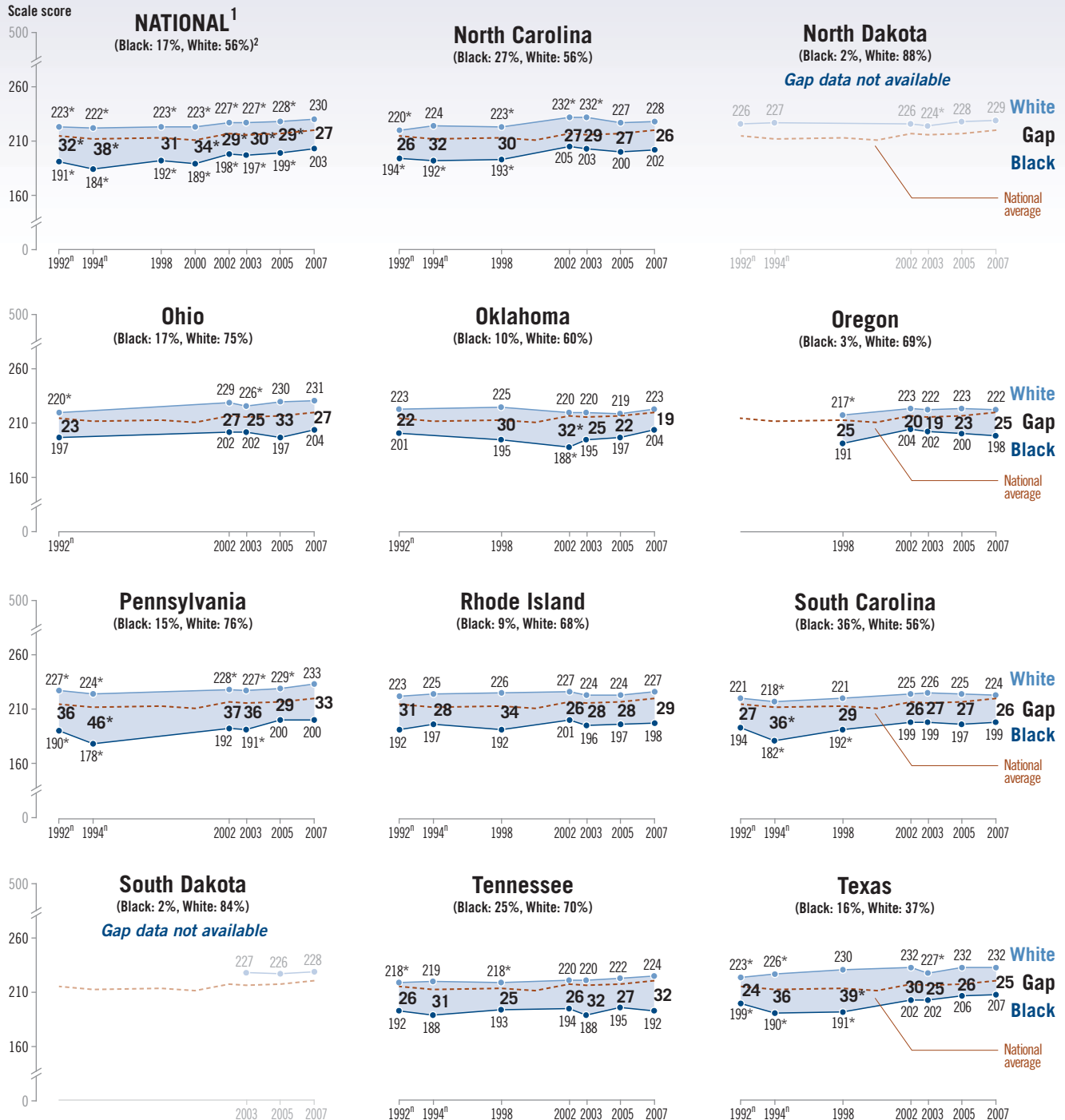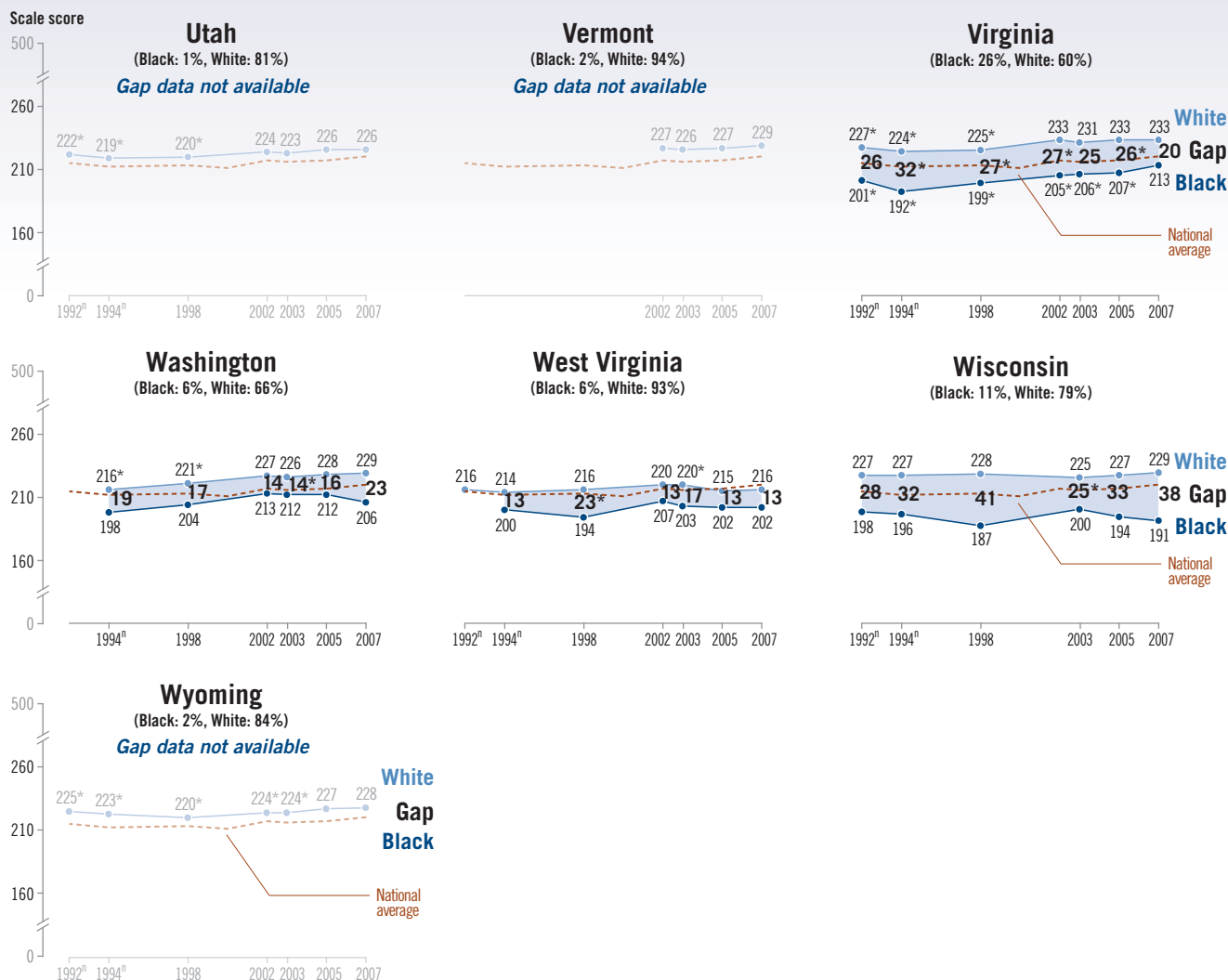

Utah
(Black: 1%, White: 81%)
Gap data not available

Vermont
(Black: 2%, White: 94%)
Gap data not available

Virginia
(Black: 26%, White: 60%)

Washington
(Black: 6%, White: 66%)

West Virginia
(Black: 6%, White: 93%)

Wisconsin
(Black: 11%, White: 79%)

Wyoming
(Black: 2%, White: 84%)
Gap data not available

[n] Accommodations were not permitted for this assessment.

* Significantly different (p<.05) from 2007.

[1] National results for assessments prior to 2002 are based on the national sample, not on aggregated state samples.

[2] Black and White percentages are based on students tested in 2007.

[3] Department of Defense Education Activity (overseas and domestic schools). Before 2005, DoDEA overseas and domestic schools were separate jurisdictions in NAEP. Pre-2005 data presented here were recalculated for comparability.

NOTE: Detail may not sum to totals due to rounding. Where data are not present, the jurisdiction did not participate or did not meet the minimum participation guidelines for reporting. State-level data were not collected in 2000. Comparative performance results may be affected by changes in exclusion rates for students with disabilities and English language learners in the NAEP samples.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1992–2007 Reading Assessments.
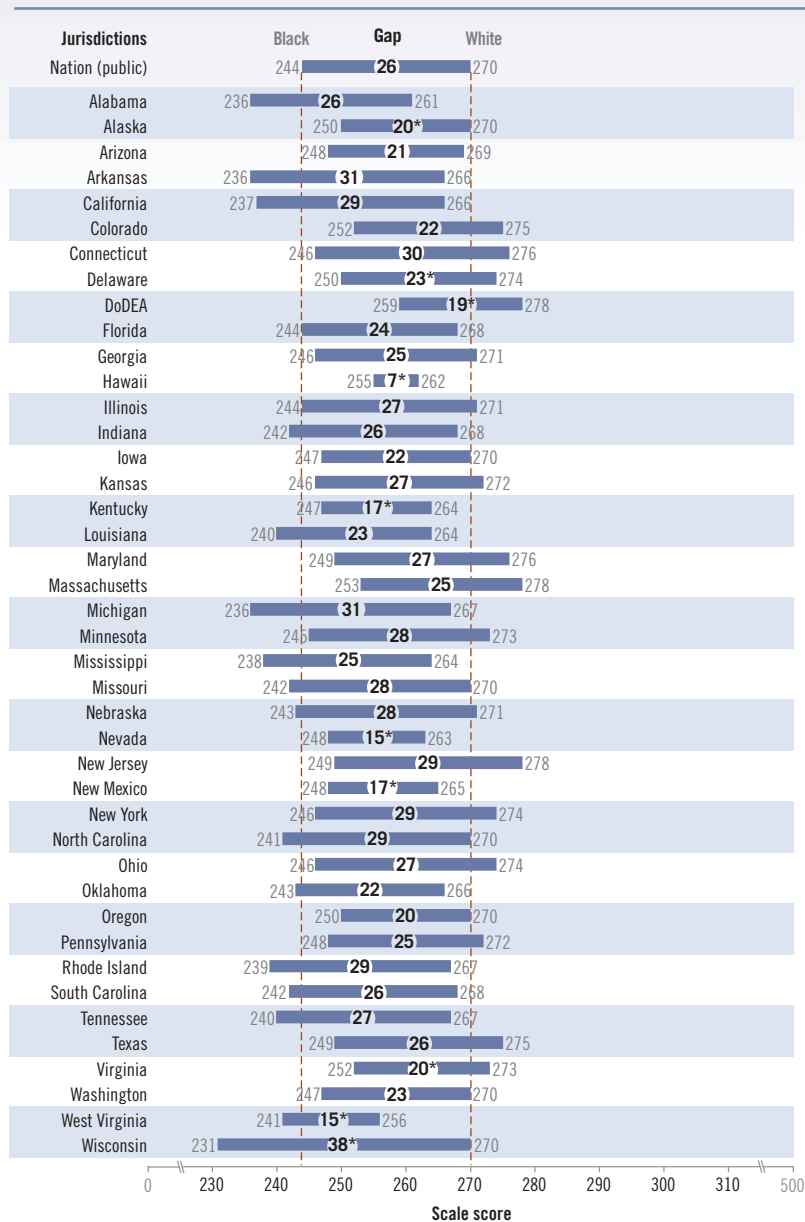
State and national reading achievement gaps at grade 8, 2007

Nine states had a Black-White gap that was smaller than the nation's 26-point gap in 2007 (Alaska, Delaware, DoDEA, Hawaii, Kentucky, Nevada, New Mexico, Virginia, and West Virginia) and one had a gap that was larger (Wisconsin). In 32 states, the gap was not significantly different from the nation's. State gaps that are either significantly larger or smaller than the national gap are indicated with asterisks (figure 23).

In Hawaii, the 7-point difference between the average scores for Black and White students was not statistically significant, and thus there was no Black-White gap for grade 8 reading in that state in 2007. In the other 41 states for which reliable data could be reported, the differences were statistically significant. The gaps ranged from 15 points in West Virginia and Nevada to 38 points in Wisconsin.

Figure 23. The Black-White achievement score gap in reading for public school students at grade 8, by state or jurisdiction: 2007

Jurisdictions	Black	Gap	White
Nation (public)	244	(26)	270
Alabama	236	(26)	261
Alaska	250	(20*)	270
Arizona	248	(21)	269
Arkansas	236	(31)	266
California	237	(29)	266
Colorado	252	(22)	275
Connecticut	246	(30)	276
Delaware	250	(23*)	274
DoDEA	259	(19*)	278
Florida	244	(24)	268
Georgia	246	(25)	271
Hawaii	255	7*	262
Illinois	244	(27)	271
Indiana	242	(26)	268
Iowa	247	(22)	270
Kansas	246	(27)	272
Kentucky	247	(17*)	264
Louisiana	240	(23)	264
Maryland	249	(27)	276
Massachusetts	253	(25)	278
Michigan	236	(31)	267
Minnesota	245	(28)	273
Mississippi	238	(25)	264
Missouri	242	(28)	270
Nebraska	243	(28)	271
Nevada	248	(15*)	263
New Jersey	249	(29)	278
New Mexico	248	(17*)	265
New York	246	(29)	274
North Carolina	241	(29)	270
Ohio	246	(27)	274
Oklahoma	243	(22)	266
Oregon	250	(20)	270
Pennsylvania	248	(25)	272
Rhode Island	239	(29)	267
South Carolina	242	(26)	268
Tennessee	240	(27)	267
Texas	249	(26)	275
Virginia	252	(20*)	273
Washington	247	(23)	270
West Virginia	241	(15*)	256
Wisconsin	231	(38*)	270

Scale score

* Significantly different (p<.05) from the nation (public) when comparing one state to the nation at a time.
[1] Department of Defense Education Activity (overseas and domestic schools).
NOTE: States whose Black or White student population size was insufficient for comparison are omitted. Reporting standards not met for District of Columbia, Idaho, Maine, Montana, New Hampshire, North Dakota, South Dakota, Utah, Vermont, and Wyoming.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessment.

43

Trends in state reading achievement gaps at grade 8, 1998–2007

Reading scores for the nation's public school students in the eighth grade were higher in 2007 than in 1992 for both Black and White students, but were not significantly different than in 1998 (figure 24, National results). Moreover, the national eighth-grade reading gap has not changed since either 1992 or 1998. NAEP first conducted eighth-grade reading assessments at the state level in 1998.
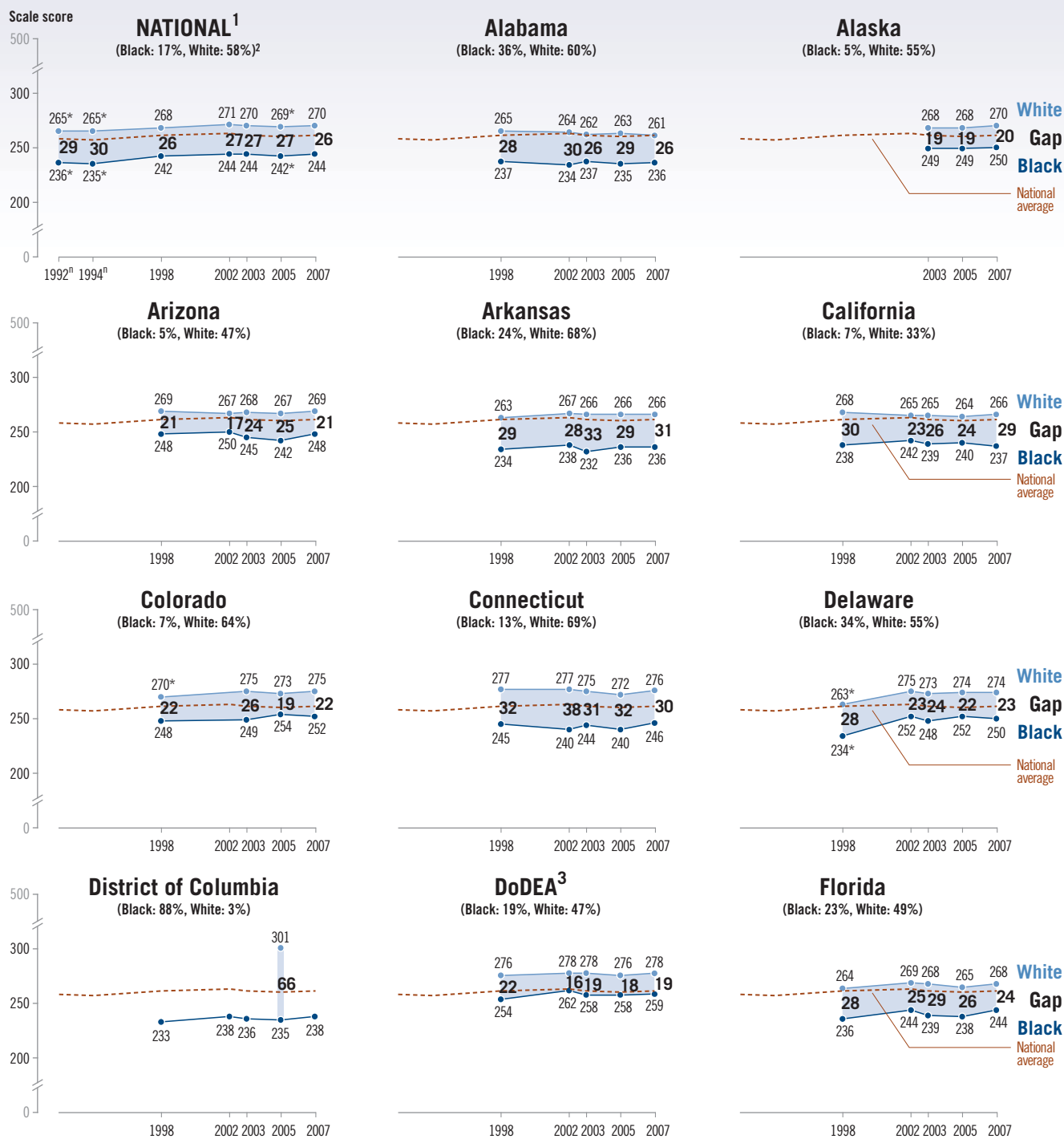
From 1998 to 2007, the Black-White score gap did not change for any state. In Delaware, scores for both Black and White eighth-graders were higher in 2007 than in 1998, but there was no significant change in gap. During this period, five other states showed significant changes in average scores in one, but not both, of the two student groups. From 2005 to 2007, the gap did not change in any state.

Despite the fact that no statistically significant changes in state gaps were identified, the 7-point difference in Black and White students' scores in Hawaii for 2007 was itself not statistically significant, so that no Black-White score gap in grade 8 reading existed for that state.
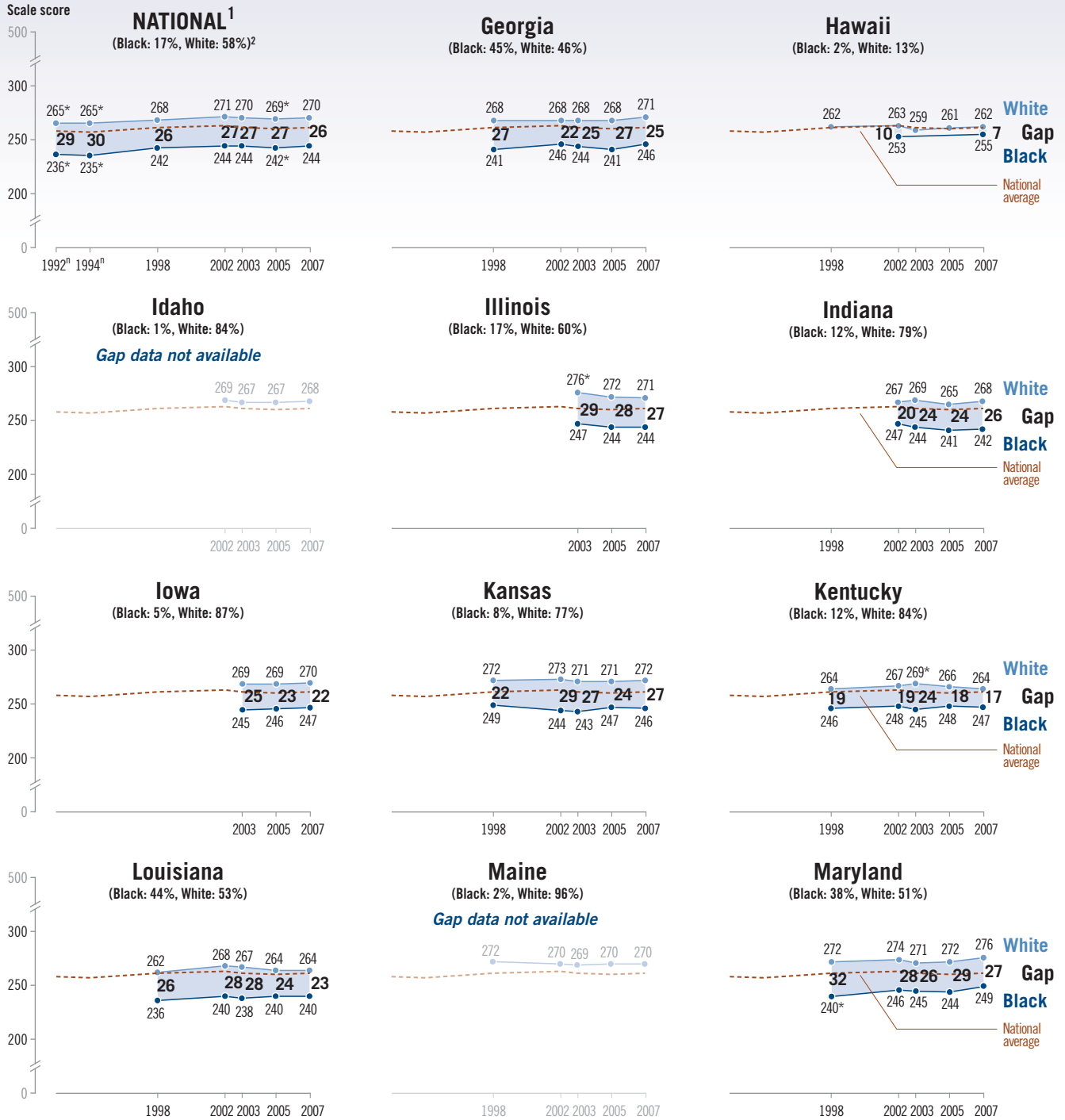
Figure 24.   Gaps in average reading scores between Black and White public school students at grade 8, by state: Various years, 1998–2007
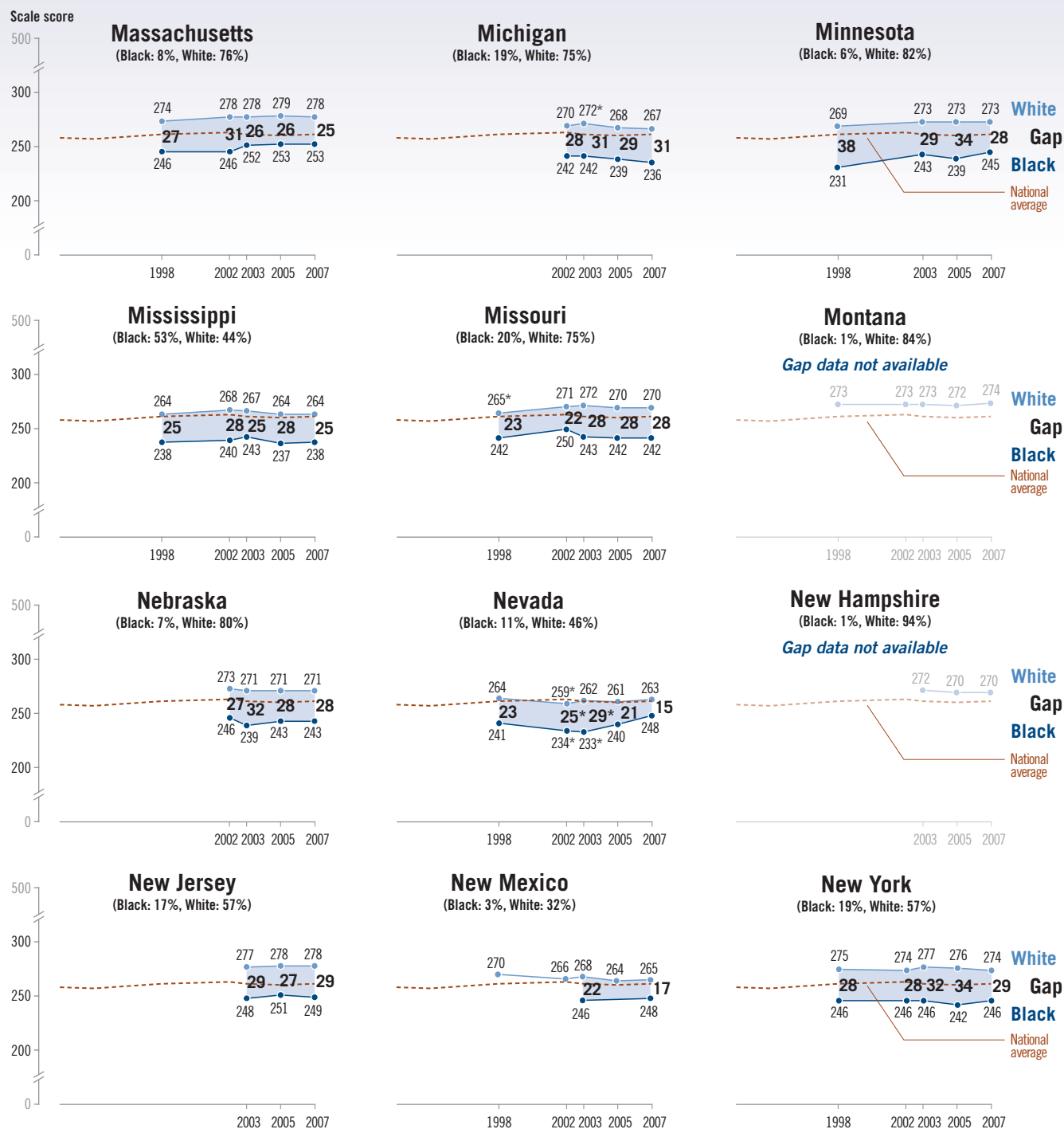

Scale score

NATIONAL[1]
(Black: 17%, White: 58%)[2]

Alabama
(Black: 36%, White: 60%)

Alaska
(Black: 5%, White: 55%)

Arizona
(Black: 5%, White: 47%)

Arkansas
(Black: 24%, White: 68%)

California
(Black: 7%, White: 33%)

Colorado
(Black: 7%, White: 64%)

Connecticut
(Black: 13%, White: 69%)

Delaware
(Black: 34%, White: 55%)

District of Columbia
(Black: 88%, White: 3%)

DoDEA[3]
(Black: 19%, White: 47%)

Florida
(Black: 23%, White: 49%)

See notes at end of figure.

45

Figure 24.   Gaps in average reading scores between Black and White public school students at grade 8, by state: Various years, 1998–2007—Continued

See notes at end of figure.

Figure 24. Gaps in average reading scores between Black and White public school students at grade 8, by state: Various years, 1998–2007—Continued

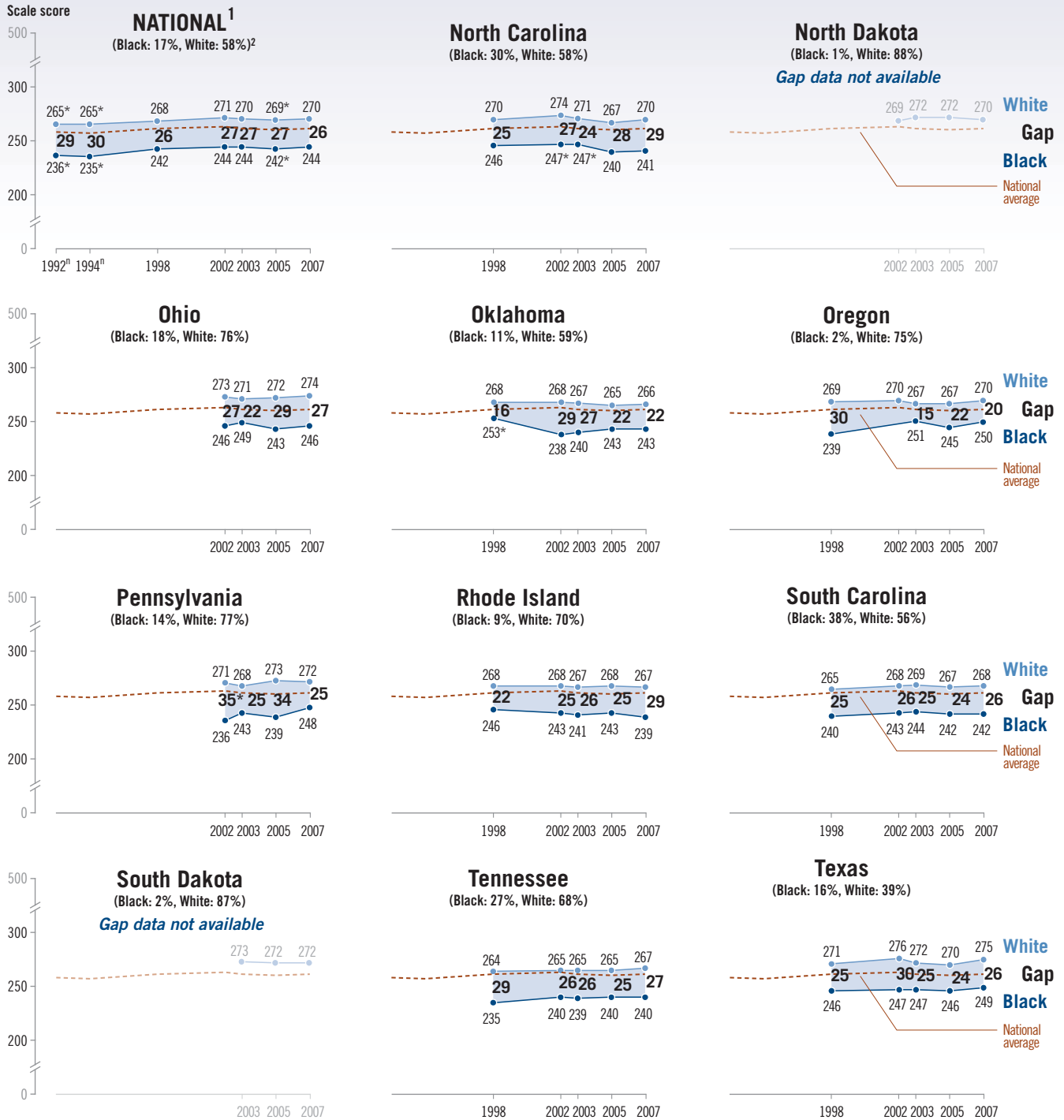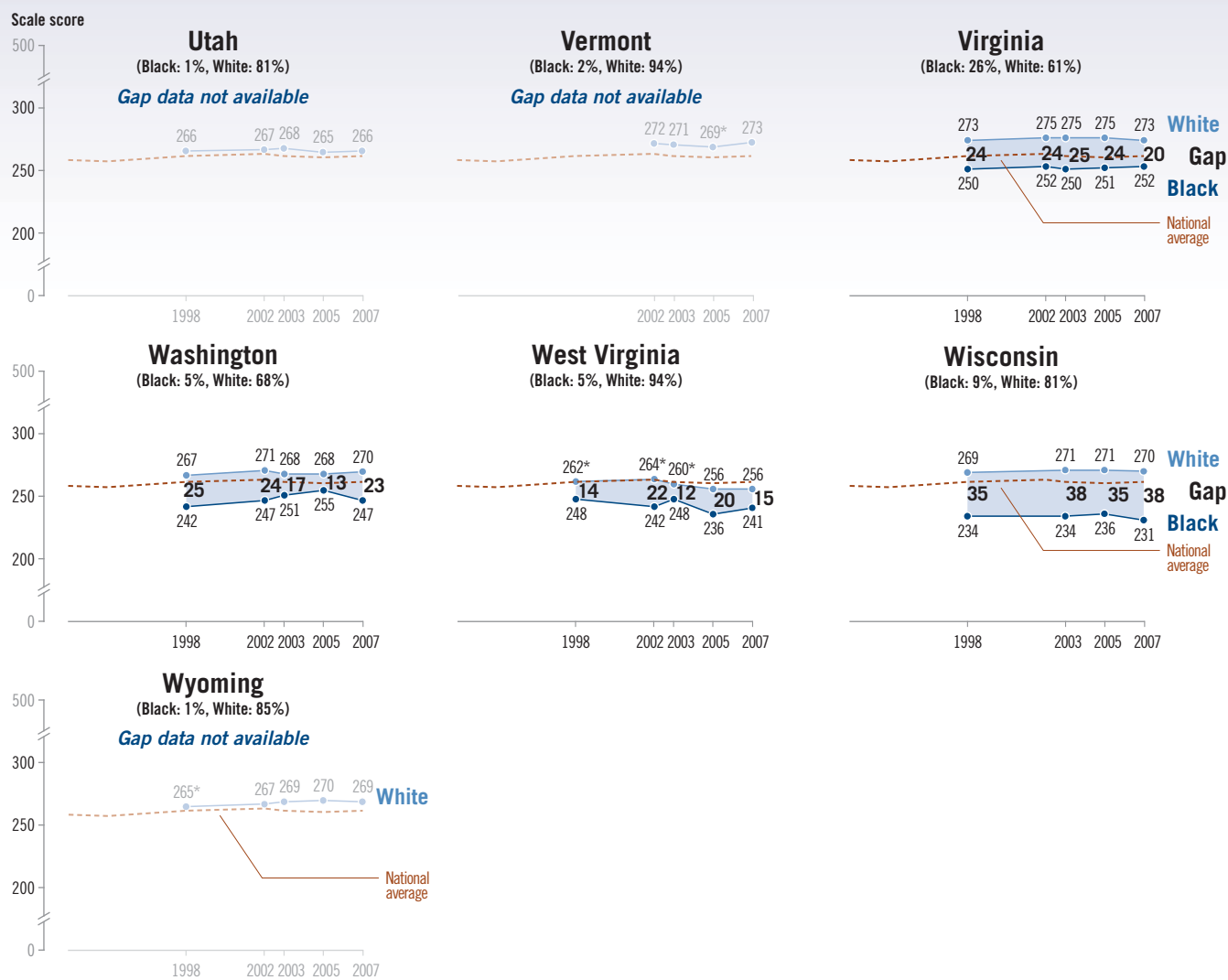Scale score

Massachusetts
(Black: 8%, White: 76%)

274 · 278 278 279 278 · White
27 · 31 26 26 25 · Gap
246 · 246 252 253 253 · Black
National average

1998 2002 2003 2005 2007

Michigan
(Black: 19%, White: 75%)

270 272* 268 267
28 31 29 31
242 242 239 236

2002 2003 2005 2007

Minnesota
(Black: 6%, White: 82%)

269 · 273 273 273 · White
38 · 29 34 28 · Gap
231 · 243 239 245 · Black
National average

1998 2003 2005 2007

Mississippi
(Black: 53%, White: 44%)

264 · 268 267 264 264
25 · 28 25 28 25
238 · 240 243 237 238

1998 2002 2003 2005 2007

Missouri
(Black: 20%, White: 75%)

265* · 271 272 270 270
23 · 22 28 28 28
242 · 250 243 242 242

1998 2002 2003 2005 2007

Montana
(Black: 1%, White: 84%)

Gap data not available

273 · 273 273 272 274 · White
Gap
Black
National average

1998 2002 2003 2005 2007

Nebraska
(Black: 7%, White: 80%)

273 271 271 271
27 32 28 28
246 239 243 243

2002 2003 2005 2007

Nevada
(Black: 11%, White: 46%)

264 · 259* 262 261 263
23 · 25* 29* 21 15
241 · 234* 233* 240 248

1998 2002 2003 2005 2007

New Hampshire
(Black: 1%, White: 94%)

Gap data not available

272 270 270 · White
Gap
Black
National average

2003 2005 2007

New Jersey
(Black: 17%, White: 57%)

277 278 278
29 27 29
248 251 249

2003 2005 2007

New Mexico
(Black: 3%, White: 32%)

270 · 266 268 264 265
22 · 17
246 · 248

1998 2002 2003 2005 2007

New York
(Black: 19%, White: 57%)

275 · 274 277 276 274 · White
28 · 28 32 34 29 · Gap
246 · 246 246 242 246 · Black
National average

1998 2002 2003 2005 2007

See notes at end of figure.

47

Figure 24. Gaps in average reading scores between Black and White public school students at grade 8, by state: Various years, 1998–2007—Continued

Scale score

NATIONAL[1]
(Black: 17%, White: 58%)[2]

North Carolina
(Black: 30%, White: 58%)

North Dakota
(Black: 1%, White: 88%)
Gap data not available

Ohio
(Black: 18%, White: 76%)

Oklahoma
(Black: 11%, White: 59%)

Oregon
(Black: 2%, White: 75%)

Pennsylvania
(Black: 14%, White: 77%)

Rhode Island
(Black: 9%, White: 70%)

South Carolina
(Black: 38%, White: 56%)

South Dakota
(Black: 2%, White: 87%)
Gap data not available

Tennessee
(Black: 27%, White: 68%)

Texas
(Black: 16%, White: 39%)

See notes at end of figure.

Figure 24. Gaps in average reading scores between Black and White public school students at grade 8, by state: Various years, 1998–2007—Continued

Scale score

Utah
(Black: 1%, White: 81%)

Gap data not available

266   267 268   265   266

1998   2002 2003 2005   2007

Vermont
(Black: 2%, White: 94%)

Gap data not available

272 271   269*   273

2002 2003 2005   2007

Virginia
(Black: 26%, White: 61%)

273   275 275 275   273   **White**
24 ··· **24 25 24 20**   **Gap**
250   252 250 251   252   **Black**

National average

1998   2002 2003 2005   2007

Washington
(Black: 5%, White: 68%)

267   271 268   268   270   **White**
25   **24 17** ·· **13** ·· **23**   **Gap**
242   247 251   255   247   **Black**

National average

1998   2002 2003 2005   2007

West Virginia
(Black: 5%, White: 94%)

262*   264* 260*   256   256
14   **22 12**   **20**   **15**
248   242 248   236   241

1998   2002 2003 2005   2007

Wisconsin
(Black: 9%, White: 81%)

269   271 271   270   **White**
35   **38 35**   38   **Gap**
234   234 236   231   **Black**

National average

1998   2003 2005   2007

Wyoming
(Black: 1%, White: 85%)

Gap data not available

265*   267 269   270   269   **White**

National average

1998   2002 2003 2005   2007

---

[n] Accommodations were not permitted for this assessment.
* Significantly different (p<.05) from 2007.
[1] National results for assessments prior to 2002 are based on the national sample, not on aggregated state samples.
[2] Black and White percentages are based on students tested in 2007.
[3] Department of Defense Education Activity (overseas and domestic schools). Before 2005, DoDEA overseas and domestic schools were separate jurisdictions in NAEP. Pre-2005 data presented here were recalculated for comparability.
NOTE: Detail may not sum to totals due to rounding. Where data are not present, the jurisdiction did not participate or did not meet the minimum participation guidelines for reporting. State-level data were not collected in 1992, 1994, or 2000. Comparative performance results may be affected by changes in exclusion rates for students with disabilities and English language learners in the NAEP samples.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1992–2007 Reading Assessments.

Appendix A: Technical Notes

This report presents data from two different assessment series, the NAEP long-term trend assessments and the main NAEP assessments. In most but not all cases, the two assessments used different procedures. Whenever a topic requires separate treatment of the two assessments, discussion of the long-term trend assessments, which present national results only, appears first, followed by the discussion of the main NAEP assessments, which present both national and state results. Discussion of main NAEP grade 12 assessments is omitted in this report because these assessments are conducted at the national level only.

Frameworks, development, administration, scoring, and analysis

Long-term trend
Overviews of these topics and more extensive information about other topics for the long-term trend assessments can be obtained from *NAEP 2004 Trends in Academic Progress*, available from the NAEP website http://nces.ed.gov/nationsreportcard/ltt/, which also provides links to earlier reports in the long-term trend series. (In 2004, the long-trend assessments in reading and mathematics were conducted for two different "studies": the "bridge study," which was identical to previous long-term assessments, and the "modified study," which will be used in future long-term assessments. The results for the 2004 assessment, reported in NAEP 2004 Trends in Academic Progress and in this report, were drawn from the bridge study.)

Main NAEP 2007 reading and mathematics assessments
For overviews of these topics, and for more extensive information about other topics for the 2007 main NAEP reading and mathematics assessments, consult the information available online at http://nces.ed.gov/nationsreportcard/reading/ and http://nces.ed.gov/nationsreportcard/mathematics/

Sources of the data

Long-term trend
This report presents national data from the 1978, 1982, 1986, 1990, 1992, 1996, 1999, and 2004 long-term trend mathematics assessments and the 1980, 1984, 1990, 1992, 1994,

1996, 1999, and 2004 long-term trend reading assessments for Black and White public school students ages 9 and 13. Earlier long-term trend assessment results are available, but only for both public and private school students combined.

Main NAEP
This report presents national data from the 1990, 1992, 1996, 2000, 2003, 2005, and 2007 main NAEP mathematics assessments and the 1992, 1994, 1998, 2002, 2003, 2005, and 2007 main NAEP reading assessments for Black and White public school students in the fourth and eighth grades. In 2000, the reading assessment was also administered in the fourth grade (see tables B-1 and B-3 in appendix B).

This report presents state data from the 1992, 1996, 2000, 2003, 2005, and 2007 fourth-grade main NAEP mathematics assessments and from the 1990, 1992, 1996, 2000, 2003, 2005, and 2007 eighth-grade main NAEP mathematics assessments, for public school students only. The main NAEP reading assessment was administered at the state level to fourth-grade public school students in 1992, 1994, 1998, 2002, 2003, 2005, and 2007 and to eighth-grade public school students in 1998, 2002, 2003, 2005, and 2007.

Nationally in 2007, Black students constituted 17 percent of the public school fourth-grade population (based on data from the NAEP reading assessment) while White students constituted 56 percent. Results for the eighth-grade were similar: 17 percent and 58 percent, respectively. However, percentages vary widely between states. For example, Black students constituted a majority of the fourth-grade population in two states, the District of Columbia (84 percent in mathematics and 86 percent in reading) and Mississippi (52 percent in mathematics and 51 percent in reading). In contrast, Black students constituted only 2 percent of the fourth-grade public school population in states such as Wyoming and South Dakota. Eighth-grade data show a similar pattern. In some cases, the Black or White student population is so small that valid data cannot be obtained.

NAEP sampling procedures

Long-term trend
The populations sampled for the 2004 NAEP long-term trend assessment results presented in this report consisted

of 9- and 13-year-old students enrolled in public elementary and secondary schools nationwide. Eligibility for the age 9 and age 13 samples was based on calendar year: students in the age 9 sample were 9 years old on January 1, 2004, with birth months January 1994 through December 1994, and students in the age 13 sample were 13 years old on January 1, 2004, with birth months January 1990 through December 1990.

Consistent with past national long-term trend assessments, students were selected for participation based on a stratified three-stage sampling plan. In the first stage, geographic primary sampling units (PSUs) were defined and selected. In the second stage, schools were selected within PSUs. In the third stage, eligible students were selected within schools. Stratification occurred at both the school level and the PSU level. A full description of the sampling plan is beyond the scope of this appendix; for additional details regarding the design and structure of the 2004 trend assessment samples, the reader should refer to the technical documentation section of the NAEP website (http://nces.ed.gov/nationsreportcard/ltt).

The first-stage sampling units, PSUs, were drawn from a list—a sampling frame—developed using the metropolitan area designations of the U.S. Census Bureau. Each NAEP PSU in the frame was intended to encompass one county or contiguous multiple counties, generally not crossing state boundaries, and contained a minimum number of school-aged children—10,000 to 15,000, depending on the region of the country.

All PSUs containing more than 800,000 students (17 in all) were automatically included in the sample. Sixty additional PSUs were selected in a non-random manner, taking into account region of the country, status as either metropolitan or non-metropolitan, percentages of racial/ethnic groups, income levels, education levels in the population, and percentage of renters, with adjustments made to compensate for the non-random manner of selection.

In the second stage of sampling, schools were sampled from within the selected PSUs. Schools were selected with probability proportional to a measure of size based on the estimated number of age-eligible students in the school.

This in turn was estimated by applying population-level percentages of age-eligible students within each grade to estimated grade enrollments for each grade, and aggregating to an age-eligible total for the school.

In the third stage of sampling, students were sampled from within schools. Sampled schools were asked to list all students with the appropriate birth dates for each specified age sample. All eligible students up to a pre-specified maximum (128 for both ages 9 and 13) were then selected for the assessment. If a school selected for the age 9 or age 13 samples had 128 or fewer students, all age-eligible students were selected into the sample for that school. Otherwise, a sample of 128 age-eligible students was taken.

The actual student and school sample sizes obtained in the NAEP long-term trend reading assessments, as well as the school and student participation rates, are presented in table A-1. Sample sizes and participation rates for the long-term trend mathematics assessments were similar. Although sampled schools that refused to participate were replaced, school participation rates were computed based on the schools originally selected for participation in the assessments. The student participation rates represent the percentage of students assessed of those invited to be assessed, including those assessed in follow-up sessions when necessary. Response rates for public school students ages 9 and 13 met NCES reporting standards for all assessments.

Table A-1.   School and student participation rates, and target student population, Long-Term Trend Reading assessment, public school students only, by age: 2004

Participation and target population	Age 9	Age 13
School participation		
Weighted school percentage	88	85
Total number of schools that participated	250	230
Student participation		
Weighted student percentage	94	92
Total number of students who participated	3,800	4,000
Target population	3,700,000	3,690,000

NOTE: The numbers of schools are rounded to the nearest ten, the numbers of students are rounded to the nearest hundred, and the numbers for target populations are rounded to the nearest ten thousand. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2004 Long-Term Trend Reading Assessments.

Main NAEP

The schools and students participating in NAEP assessments are chosen to be nationally representative. Samples of schools and students are selected from each state and from the District of Columbia and Department of Defense Education Activity (DoDEA) schools. The results from the assessed students are combined to provide accurate estimates of overall national performance and of the performance of individual states.

NCES has changed the main NAEP sampling methods over the years. From 1990 through 2000, the national sample was collected separately from the state samples. The 2002 national sample was the sum of all the state samples of the participating states, plus small samples from the few states that did not participate. In 2003, 2005, and 2007, all states participated and the national sample was the aggregate of the samples from all states, the District of Columbia, and the DoDEA schools. The main NAEP national samples in reading and mathematics since 2002 have been larger than in previous assessment years. Thus, smaller score differences between years or between types of student were found to be statistically significant than would have been detected in previous assessments.

From 1990 through 2001, NCES oversampled schools with high minority populations (Black and Hispanic) in the national sample. Beginning in 2002, this practice was discontinued because the state samples were large enough to ensure adequate coverage for these populations. Prior to 2002, NAEP results were weighted to compensate for the oversampling.

In 2003, 2005, and 2007, results were weighted to take into account the fact that states, and schools within states, represent different proportions of the overall national population. For example, since the number of students assessed in most states is roughly the same (to allow for stable state estimates and administrative efficiencies), the results for students in less populous states are assigned smaller weights than the results for students in more populous states. Sampling weights are also used to account for lower sampling rates for very small schools and are used to adjust for school and student nonresponse.

NAEP samples for reading and mathematics assessments administered from 1990 through 2007 are discussed in more detail below.

The NAEP 2007 mathematics and reading assessments were administered to fourth- and eighth-graders in all states. This report includes data for public school students for both the nation and all states. All 50 states, the District of Columbia, and the DoDEA schools met the minimum guidelines for reporting their results in 2007 for both assessments.

In order to obtain a representative sample for reporting national and state public school results in 2007, NCES sampled and assessed approximately 183,000 fourth-graders from 7,300 schools and 155,000 eighth-graders from 6,400 schools for the reading assessment and approximately 190,000 fourth-graders from 7,300 schools and 147,000 eighth-graders from 6,400 schools for the mathematics assessment.

Each selected school that participated in the assessment and each student assessed represent a portion of the population. The schools were selected out of approximately 51,000 fourth-grade and 27, 000 eighth-grade public schools. The students selected from these schools represented the total population of approximately 3.4 million fourth-grade and 3.6 million eighth-grade public school students. These totals include the public schools in the 50 states and the District of Columbia.

Schools in the DoDEA school system are classified as "nonpublic" by NCES and their results are not included in the determination of NAEP national public average scale scores. These schools are not "private" because they are operated by the federal government and they are not "public" because only children of U.S. military personnel can attend them. For comparison purposes, the system is treated as a state and results are compared with the scores of the 50 states and the District of Columbia.

Table A-2 provides a summary of the 2007 national and state school and student participation rates for the reading grade 8 assessment sample. Rates for reading grade 4 and mathematics grades 4 and 8 in 2007 were similar, as

were the rates for the 2003 and 2005 assessments. Readers who want more detail should consult the 2007, 2005 and 2003 report cards, available online at http://nces.ed.gov/pubsearch/getpubcats.asp?sid=031.

Participation rates in table A-2 are presented for public schools and public school students in grade 8 reading. The school participation rate is a school-centered, weighted percentage of schools participating in the assessment. This rate is based only on the schools that were initially selected for the assessment. The numerator of this rate is the estimated number of schools represented by the initially selected schools that participated in the assessment. The denominator is the estimated number of schools represented by the initially selected schools that had eligible students enrolled.

Also presented in table A-2 are weighted student participation rates. The numerator of this rate is the estimated number of students who are represented by the students assessed (in either an initial session or a makeup session). The denominator of this rate is the estimated number of students represented by the eligible sampled students in participating schools.

The term "eligible students" used in the two preceding paragraphs refers to students who can meaningfully participate in NAEP. Students excluded from NAEP assessments on the grounds that they cannot meaningfully participate—whether students with disabilities or English language learners—are not part of the population of interest. Initially selected schools that had no eligible students enrolled are excluded from the denominator of the school participation rate because they contained no students who were part of the population of interest. For similar reasons, the denominator of the weighted student participation rate consists only of eligible sampled students.

The fourth column gives the number of public school students who were assessed in each of the jurisdictions. The final column of table A-2 gives the target populations for each jurisdiction, that is, the eighth-grade population for that jurisdiction.

The national target population per grade for all main NAEP assessments 1990–2007 ranged from about 3.25 million

to about 3.75 million. In the 1990–1996 assessments, the number of schools sampled per assessment and grade for the national sample ranged from approximately 120 to 230, while the number of students assessed ranged from approximately 5,200 to 9,900. In the 1998–2000 assessments, the number of schools sampled per assessment and grade ranged from approximately 330 to 390, while the number of students assessed ranged from approximately 6,100 to 9,000.

The state target populations for all main NAEP assessments 1990–2007 ranged from approximately 5,000 in the District of Columbia and 9,000 in sparsely populated states like Wyoming and Alaska to approximately 450,000 in California, followed by approximately 325,000 in Texas.

In the 1990–2000 state assessments, the number of schools sampled per assessment and grade ranged from approximately 30 to 150, while the number of students assessed ranged from approximately 1,000 to 5,900. In the 2003–2007 state assessments, the number of schools sampled per assessment and grade ranged from approximately 40 to 250, while the number of students assessed ranged from approximately 1,700 to 10,700.

In earlier NAEP assessments, NCES would select substitute schools that would be used to augment the original sample if a large number of schools from the sample failed to participate. School and student participation rates were given both before and after substitution. Because the No Child Left Behind Act requires states to participate in the main NAEP reading and mathematics assessments at the fourth and eighth grades in order to qualify for full Title I education funding, participation rates are very high and NCES no longer selects substitute schools for these assessments.

In order to ensure unbiased samples, NCES and the National Assessment Governing Board, which establishes policy for NAEP, set minimums for the school participation rate before substitution of replacement schools for any sample. From 1990 through 2002, the standard for the state assessments required that the weighted school participation rate before substitution of replacement schools

Table A-2. School and student participation rates, and target student population, grade 8 reading assessment, public school students only, by state or jurisdiction: 2007

Jurisdiction	School participation		Student participation		
	Weighted school percentage	Total number of schools that participated	Weighted student percentage	Total number of students who participated	Target population
Nation (public)	**100**	**6,410**	**92**	**154,700**	**3,558,000**
Alabama	100	120	93	2,800	56,000
Alaska	99	110	91	2,600	9,000
Arizona	100	130	90	2,800	73,000
Arkansas	100	120	93	2,500	34,000
California	100	310	92	8,600	477,000
Colorado	98	120	92	2,800	57,000
Connecticut	97	100	92	2,700	42,000
Delaware	100	50	93	2,800	10,000
District of Columbia	100	50	88	1,800	5,000
DoDEA[1]	98	60	94	1,700	5,000
Florida	100	160	91	4,100	193,000
Georgia	100	120	93	3,500	120,000
Hawaii	100	70	91	2,800	13,000
Idaho	99	110	93	2,900	20,000
Illinois	100	200	93	4,000	150,000
Indiana	100	110	92	2,700	80,000
Iowa	100	130	93	2,800	36,000
Kansas	100	150	94	2,800	34,000
Kentucky	100	110	93	2,600	46,000
Louisiana	100	110	92	2,400	47,000
Maine	98	130	93	2,700	15,000
Maryland	100	110	90	2,700	64,000
Massachusetts	100	140	93	3,600	70,000
Michigan	100	120	91	2,600	119,000
Minnesota	99	140	92	3,000	62,000
Mississippi	100	110	93	2,700	36,000
Missouri	100	130	92	2,900	70,000
Montana	98	170	92	2,600	11,000
Nebraska	100	120	94	2,700	21,000
Nevada	100	70	88	2,600	28,000
New Hampshire	98	90	92	2,900	16,000
New Jersey	97	110	92	2,800	104,000
New Mexico	100	110	89	2,600	25,000
New York	100	160	90	3,800	206,000
North Carolina	100	150	91	4,300	104,000
North Dakota	98	190	95	2,200	8,000
Ohio	100	190	92	3,500	135,000
Oklahoma	100	150	92	2,600	42,000
Oregon	100	110	92	2,700	39,000
Pennsylvania	100	110	92	2,800	140,000
Rhode Island	100	60	92	2,800	12,000
South Carolina	100	110	94	2,700	52,000
South Dakota	99	140	95	2,800	10,000
Tennessee	100	120	92	2,800	74,000
Texas	100	220	92	7,100	294,000
Utah	100	100	91	2,800	36,000
Vermont	100	120	93	2,000	7,000
Virginia	100	110	93	2,800	91,000
Washington	100	130	91	3,000	78,000
West Virginia	100	120	92	2,900	21,000
Wisconsin	98	130	92	2,700	62,000
Wyoming	100	80	92	2,000	7,000

[1] Department of Defense Education Activity (overseas and domestic schools).
NOTE: The numbers of schools are rounded to the nearest ten, the numbers of students are rounded to the nearest hundred, and the target population is rounded to the nearest thousand. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessment.

be 70 percent or higher. Beginning in 2003, the standard was raised to 85 percent. All data presented in this report are based on samples meeting the standards in effect at the time of the assessment.

Since 1990, the national weighted public school participation rate before substitution for the grade 4 and 8 reading and mathematics assessments has ranged from 76 percent to 100 percent. Prior to 2003, a few states did not meet the 70 percent standard. From 1990 through 2002, the weighted public school participation rate before substitution for states whose results are reported here ranged from 70 percent to 100 percent.

For more information on all the NAEP assessments referenced in this report, consult the individual reports devoted to them, available from the NCES website at http://nces.ed.gov/pubsearch/getpubcats.asp?sid=031.

Understanding NAEP reporting groups

NAEP results are provided for groups of students defined by shared characteristics—race/ethnicity, eligibility for free/reduced-price school lunch, and gender, for example.

Based on participation rate criteria, results are reported for groups only when sufficient numbers of students and adequate school representation are present. The minimum requirement is a total of at least 62 students in a particular group, assessed in at least five different locations. However, the data for all students, regardless of whether their group was reported separately, were included in computing over-all national results. Definitions of the student groups discussed in this report follow.

Race/ethnicity

Long-term trend

In long-term trend NAEP, data about student race/ethnicity is based on the assessment administrator's observation. Self-reported race/ethnicity data has been collected since 1984, and school records-based race/ethnicity data has been collected starting in 2004, but all long-term trend results are reported based on observed race/ethnicity.

Main NAEP

In all main NAEP assessments, data about student race/ethnicity are collected from two sources: school records and student self-reports. In this report, the race/ethnicity variable has been based on the race reported by the school for all assessment years. In the rare cases when school-recorded information is missing, student-reported data are used to determine race/ethnicity.

Schools sampled for NAEP are asked to provide lists of all students in the target grade(s) along with basic demographic information, including race/ethnicity. Students are categorized into one of five mutually exclusive racial/ethnic categories plus "other." Administration schedules—also referred to as student rosters—are created that include the list of sampled students along with their basic demographic information. These data are checked and updated during data collection. This race/ethnicity information is available for all sampled students: those that participated and those that were absent or excluded.

All students who take a NAEP assessment complete a section of general student background questions, including questions about their race/ethnicity. Separate questions are asked about students' Hispanic ethnic background and about students' race. This race/ethnicity information is available just for students who participated in the assessment and not for those who were absent or excluded. See http://nces.ed.gov/nationsreportcard/bgquest.asp for more information.

The mutually exclusive racial/ethnic categories are White (non-Hispanic), Black (non-Hispanic), Hispanic, Asian/Pacific Islander, American Indian (including Alaska Native), and Unclassified. Unclassified students are those whose school-reported race was "other," or "unavailable," or was missing, or who self-reported more than one race category (i.e., "multi-racial") or none. Hispanic students may be of any race. Only results for White (non- Hispanic) and Black (non-Hispanic) students are contained in this report. Information based on student self-reported race/ethnicity is available on the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/nde).

Eligibility for free/reduced-price school lunch

Long-term trend

The long-term trend assessments do not report results based on school lunch eligibility.

Main NAEP

As part of the Department of Agriculture's National School Lunch Program, schools can receive cash subsidies and donated commodities in return for offering free or reduced-price lunches to eligible children. Based on available school records, students were classified as currently eligible for either free lunch or reduced-price lunch, or not eligible. Eligibility for the program is determined by a student's family income in relation to the federally established poverty level. Free lunch qualification is set at 130 percent of the poverty level or below, and reduced-price lunch qualification is set at between 130 and 185 percent of the poverty level. (For the period July 1, 2006, through June 30, 2007, for a family of four, 130 percent of the poverty level was $26,000, and 185 percent was $37,000. See http://www.fns.usda.gov/cnd/lunch for more information.) The classification applies only to the school year when the assessment was administered and is not based on eligibility in previous years. If school records were not available, the student was classified as "Information not available." If the school did not participate in the program, all students in that school were classified as "Information not available." As a result of improvements in the quality of the data on students' eligibility for NSLP, the percentage of students for whom information was not available has decreased in comparison to the percentages reported prior to the 2003 assessment. Therefore, trend comparisons are only made back to 2003 in this report.

Gender

Both long-term trend and NAEP assessments identify students as male or female based on school records.

Inclusion and exclusion

Long-term trend

Some students selected for participation in the NAEP long-term trend assessments were identified as English language learners (ELL) or students with disabilities (SD). In all previous long-term trend assessments, if it was decided that a student classified as SD or ELL could not meaningfully participate in the NAEP assessment for which he or she was selected, the student was, according to NAEP guidelines, excluded from the assessment.

For each student selected to participate in NAEP who was identified as either SD or ELL, a member of the school staff most knowledgeable about the student completed an SD/ELL questionnaire. Students with disabilities were excluded from the assessment if an individualized education program (IEP) team or equivalent group determined that the student could not participate in assessments such as NAEP; if the student's cognitive functioning was so severely impaired that the student could not participate; or if the student's IEP required that the student be tested with an accommodation or adaptation not permitted or available in NAEP, and the student could not demonstrate his/her knowledge of the assessment subject area without that accommodation or adaptation. A student who was identified as ELL and who was a native speaker of a language other than English was excluded if the student had received instruction in the assessment's subject area (e.g., reading or mathematics) primarily in English for less than three school years, including the current year, or if the student could not demonstrate his or her knowledge of reading or mathematics in English without an accommodation or adaptation.

Prior to 2004, NAEP long-term trend assessments did not allow accommodations for SD or ELL students. In that year, two versions of the long-term trend assessment were given, the "bridge" (unmodified) version, which did not allow accommodations, and the "modified" version, which

did. In 2004, results were only reported for the bridge assessment and all results from the 2004 Long-Trend Assessment appearing in this report are drawn from the bridge assessment. However, table A-3 presents exclusion rates for both versions of the 2004 assessment in order to give all the available information on the 2004 exclusion rates for Black and White students.

In the 2004 bridge assessment, and in all prior administrations of the long-term trend assessment, student race/ethnicity was determined by NCES contractor staff administering the assessment in the individual classrooms. These staff never met the excluded students, so no records of the race/ethnicity of excluded students were kept.

In contrast, the 2004 modified assessment determined student race/ethnicity by using school records, which did provide information on the race/ethnicity of excluded students. Exclusion data from the 2004 modified assessment are provided here to provide information on 2004 exclusion rates for Black and White students, even though this report does not include student achievement data drawn from the modified assessment.

Table A-3. National Long-Term Trend mathematics and reading exclusion rates as percentages of the total sample, by age, type of assessment and race/ethnicity: 2004

	Age 9		Age 13	
	Bridge	Modified	Bridge	Modified
Mathematics				
Total	8	3	9	3
White	†	2	†	3
Black	†	4	†	4
Reading				
Total	9	6	9	5
White	†	4	†	5
Black	†	4	†	6

† Not applicable.
NOTE: The 2004 bridge assessment, and all previous administrations of the long-term trend assessment, did not obtain information on the race/ethnicity of excluded students.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2004 Long-Term Trend Mathematics and Reading Assessments.

Main NAEP

The NAEP program has always endeavored to assess all students selected as a part of its sampling process. In all NAEP schools, accommodations will be provided as necessary for students with disabilities (SD) and/or English language learners (ELL) or limited English proficient (LEP) students. (ELL is the term used since the NAEP 2005 reports; LEP was used before 2005.) The accommodations are available to students whose Individualized Education Program (IEP) specifically requires them. Because some ELL students do not have an IEP, decisions about accommodations for these students are typically made by knowledgeable school staff.

The NAEP program has established procedures to include as many SD and ELL students as possible in the assessments. School staff make the decisions about whether to include such a student in a NAEP assessment, and which testing accommodations, if any, they should receive. The NAEP program furnishes tools to assist school personnel in making those decisions.

A sampling procedure is used to select students at each grade being tested. Students are selected on a random basis, without regard to SD or ELL status. Once the students are selected, the schools identify which have SD or ELL status. School staff who are familiar with these students are asked a series of questions to help them decide whether each student should participate in the assessment and whether the student needs accommodations.

Inclusion in NAEP of an SD or ELL student is encouraged if that student (a) participated in the regular state academic assessment in the subject being tested, and (b) if that student can participate in NAEP with the accommodations NAEP allows. Even if the student did not participate in the regular state assessment, or if he/she needs accommodations NAEP does not allow, school staff are asked whether that student could participate in NAEP with the allowable accommodations.

History of NAEP Inclusion Policy Although NAEP has always endeavored to assess as high a proportion of

sampled students as is possible, prior to 1996 NAEP did not allow accommodations for SD or ELL students. This resulted in exclusion of some students who could not meaningfully participate in the assessment without accommodations.

The passage of the Individuals with Disabilities Education Act (IDEA), as amended in 1997, led states and districts to identify increasing numbers of students as requiring accommodations in assessments in order to fairly and accurately show their abilities. It was important for NAEP to be as consistent as possible with testing practices in most states and districts while maintaining the ability to compare more recent NAEP results to those from 1990, 1992, and 1994, when accommodations were not allowed. (Accommodations were not allowed in NAEP state assessments until 1996.)

Before the 2005 assessment (when the selection process was detailed in a series of questions), guidelines were specified by NAEP. A student identified on the Administration Schedule as having a disability (SD), that is, a student with an Individualized Education Program (IEP) or equivalent classification, should be included in the NAEP assessment unless:

■ The IEP team or equivalent group had determined that the student could not participate in assessments such as NAEP, or

■ The student's cognitive functioning was so severely impaired that he or she could not participate, or

■ The student's IEP required that the student be tested with an accommodation that NAEP did not permit, and the student could not demonstrate his or her knowledge of the subject without that accommodation.

A student who was identified as LEP or ELL and who was a native speaker of a language other than English should be included in the NAEP assessment unless:

■ The student had received reading or mathematics instruction primarily in English for less than 3 school years including the current year, and

■ The student could not demonstrate his or her knowledge of the subject in English even with an accommodation permitted by NAEP.

The phrase "less than 3 school years including the current year" meant 0, 1, or 2 school years. Therefore, the guidelines below were used:

■ Include without any accommodation all LEP or ELL students who had received instruction in the subject primarily in English for 3 years or more and those who were in their third year;

■ Include without any accommodation all other such students who could demonstrate their knowledge of the subject without an accommodation;

■ Include and provide accommodations permitted by NAEP to other such students who can demonstrate their knowledge of the subject only with those accommodations; and

■ Exclude LEP or ELL students only if they could not demonstrate their knowledge of the subject even with an accommodation permitted by NAEP.

The percentages of students excluded from NAEP may vary from one state to another, as well as across years. National exclusion rates for Black and White SD and/or ELL students in 2007 may be found in table A-4. The "total" rates include all students, not just those who are Black or White. For information on state exclusion rates, see table A-5. For more information on Main NAEP inclusion and exclusion, go to http://nces.ed.gov/nationsreportcard/about/inclusion.asp

Table A-4.   National mathematics and reading exclusion rates as percentages of the total sample, public schools only, by grade and race/ethnicity: 2007

Grade and race/ethnicity	Mathematics	Reading
Grade 4		
Total	3	6
White	2	4
Black	4	7
Grade 8		
Total	4	6
White	4	4
Black	6	7

NOTE: "Total" exclusion percentages are for all public school students, not just Black and White.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics and Reading Assessments.

Table A-5. Mathematics and reading exclusion rates as percentages of the total sample, public schools only, by grade, race/ethnicity and jurisdiction: 2007

| | Percentage of students with a disability and/or English language learner, excluded in 2007 | | | | | | | |
| Jurisdiction | Grade 4 Mathematics | | Grade 8 Mathematics | | Grade 4 Reading | | Grade 8 Reading | |
	White	Black	White	Black	White	Black	White	Black
Nation (public)	2	4	3	6	4	7	4	7
Alabama	1	2	2	4	3	3	3	4
Alaska	1	2	4	6	3	3	2	2
Arizona	2	4	2	4	3	4	3	6
Arkansas	2	5	2	3	5	8	4	7
California	1	3	1	3	2	5	2	4
Colorado	2	4	1	2	2	7	2	4
Connecticut	1	1	1	1	2	3	2	4
Delaware	4	6	5	8	9	13	5	8
District of Columbia	2	5	‡	10	6	12	‡	13
DoDEA	1	1	2	1	4	5	2	3
Florida	1	4	2	3	4	6	3	5
Georgia	2	3	4	5	7	8	6	7
Hawaii	1	1	1	‡	3	1	2	0
Idaho	1	‡	1	‡	3	‡	3	‡
Illinois	3	5	4	10	5	6	4	6
Indiana	2	4	5	9	3	5	4	8
Iowa	1	2	2	6	3	13	4	15
Kansas	2	6	4	4	4	8	4	7
Kentucky	3	3	7	5	7	10	8	7
Louisiana	1	3	2	4	3	5	2	4
Maine	3	4	5	‡	6	‡	5	‡
Maryland	2	5	6	9	5	10	4	10
Massachusetts	4	6	8	12	5	9	6	10
Michigan	3	4	4	7	4	6	5	9
Minnesota	2	3	2	3	3	9	3	10
Mississippi	1	1	1	3	2	3	2	4
Missouri	3	4	5	6	4	4	3	4
Montana	2	‡	3	‡	4	‡	4	‡
Nebraska	2	5	2	3	4	8	3	2
Nevada	2	6	3	4	5	7	3	9
New Hampshire	2	8	3	13	4	13	3	‡
New Jersey	1	4	2	6	5	8	4	10
New Mexico	2	4	2	3	6	9	4	11
New York	2	2	4	4	5	5	5	4
North Carolina	1	3	1	3	1	3	2	3
North Dakota	4	‡	5	‡	7	34	8	‡
Ohio	4	8	6	12	7	12	8	10
Oklahoma	5	4	8	11	6	7	6	10
Oregon	2	5	3	5	4	7	2	5
Pennsylvania	2	3	3	4	4	6	4	8
Rhode Island	1	4	2	5	3	4	3	6
South Carolina	1	2	4	7	4	4	5	7
South Dakota	1	1	2	‡	5	17	5	‡
Tennessee	5	7	6	7	9	12	7	8
Texas	5	6	4	7	6	8	4	8
Utah	2	‡	2	‡	5	‡	4	‡
Vermont	2	‡	4	‡	6	29	5	‡
Virginia	4	6	4	9	7	9	6	9
Washington	2	3	2	7	4	6	3	10
West Virginia	1	1	2	5	2	1	2	4
Wisconsin	2	4	3	9	3	8	5	12
Wyoming	1	‡	2	‡	3	‡	3	‡

‡ Reporting standards not met.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics and Reading Assessments.

Accommodations

Long-term trend

The long-term trend results presented in this report are drawn from assessments that did not permit accommodations for students with disabilities (SD) and English language learners (ELL). Future long-term trend assessments will allow such accommodations.

Main NAEP

From 1990 through 1994 for the nation—and through 1996 for the states—main NAEP assessments did not allow accommodations for either SD or ELL students. Since then, accommodations have been permitted for those SD and ELL students who need accommodations in order to participate, unless the accommodation would change the nature of what is being tested.

To accomplish this goal, students who receive accommodations in their state's assessments are offered the same accommodations on NAEP, except where an accommodation would change the nature of what is being tested. For example, passages and questions in the reading test are not permitted to be read aloud to the student, because that accommodation would make it a test of listening instead of a test of reading. Similarly, reading passages and questions cannot be presented in a language other than English.

It should be noted that students assessed with accommodations typically received some combination of accommodations. For example, students assessed in small groups (as compared with standard NAEP sessions of about 30 students) usually received extended time. In one-on one administrations, students often received assistance in recording answers (e.g., use of a scribe or computer) and were afforded extra time.

The most common accommodations are small-group administration, extended time, one-on-one administration, the use of a scribe or computer, and the use of a bilingual book (mathematics only). See http://nces.ed.gov/nationsreportcard/tdw/instruments/accomm.asp for more details on NAEP accommodations. For state accommodation rates for SD and ELL students in 2007 see the Technical Notes sections of *The Nation's Report Card: Mathematics 2007* at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007494 and *The Nation's Report Card: Reading 2007* at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007496.

Drawing inferences from the results

The reported statistics for both long-term trend and main NAEP are estimates and are therefore subject to a measure of uncertainty. There are two sources of such uncertainty. First, NAEP uses a sample of students rather than testing all students. Second, all assessments have some amount of uncertainty related to the fact that they cannot ask all questions that might be asked in a content area. The magnitude of this uncertainty is reflected in the standard error of each of the estimates. When the percentages or average scale scores of certain groups are compared, the estimated standard error should be taken into account. Therefore, the comparisons are based on statistical tests that consider the estimated standard errors of the statistics being compared and the magnitude of the difference between the averages or percentages.

Standard errors for the NAEP scores and percentages presented in this report for both assessments are available on the NAEP website (http://nces.ed.gov/nationsreportcard/naepdata).

The differences between statistics—such as comparisons of two groups of students' average scale scores—that are discussed in this report are determined by using standard errors. Comparisons are based on statistical tests that consider both the size of the differences and the standard errors of the two statistics being compared. Estimates based on smaller groups are likely to have relatively large standard errors. As a consequence, a numerical difference that seems large may not be statistically significant.

Furthermore, differences of the same magnitude may or may not be statistically significant, depending upon the size of the standard errors of the statistics. For example, a 3-point change in the gap between Black and White fourth-graders nationwide may be significant, while a

3- point change in the gap between Black and White fourth-graders in Kansas may not be. The differences described in this report have been determined to be statistically significant at the .05 level with appropriate adjustments for part-to-whole and multiple comparisons.[1]

In the tables and figures of this report, the symbol (*) is used to indicate that a score or percentage is significantly different from another. In addition, any difference between scores or percentages that is identified as higher, lower, larger, smaller, narrower, or wider in this report, including within-group differences not marked in tables and figures, meets the requirements for statistical significance.

Weighting and variance estimation

In both long-term trend and main NAEP a complex sample design was used to select the students who were assessed. The properties of a sample selected through such a design could be very different from those of a simple random sample, in which every student in the target population has an equal chance of selection and in which the observations from different sampled students can be considered to be statistically independent of one another. Therefore, the properties of the sample for the data collection design were taken into account during the analysis of the assessment data.

One way that the properties of the sample design were addressed was by using sampling weights to account for the fact that the probabilities of selection were not identical for all students. All population and subpopulation characteristics based on the assessment data were estimated using sampling weights. These weights included adjustments for school and student nonresponse.

Not only must appropriate estimates of population characteristics be derived, but appropriate measures of the degree of uncertainty must be obtained for those statistics. Two components of uncertainty are accounted for in the variability of statistics based on student ability: (1) the uncertainty due to sampling only a relatively small number of students, and (2) the uncertainty due to sampling only a relatively small number of cognitive questions. The first component accounts for the variability associated with the estimated percentages of students who had certain background characteristics or who answered a certain cognitive question correctly.

Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable measure of uncertainty for any student information that can be observed without error. However, because each student typically responds to only a few questions within a content area, the scale score for any single student would be imprecise. In this case, NAEP's marginal estimation methodology can be used to describe the performance of groups and subgroups of students. The estimate of the variance of the students' posterior scale score distributions (which reflect the imprecision due to lack of measurement accuracy) is computed. This component of variability is then included in the standard errors of NAEP scale scores.[2]

Analyzing group differences in averages and percentages

In both long-term trend and main NAEP, statistical tests determine whether, based on the data from the groups in the sample, there is strong enough evidence to conclude that the averages or percentages are actually different for those groups in the population. If the evidence is strong (i.e., the difference is statistically significant), the report describes the group averages or percentages as being different (e.g., one group performed higher or lower than another group), regardless of whether the sample averages or percentages appear to be approximately the same. The reader is cautioned to rely on the results of the statistical tests rather than on the apparent magnitude of the difference between sample averages or percentages when

[1] Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society,* Series B, no. 1, 289–300.

[2] For further detail, see Johnson, E.G., and Rust, K.F. (1992). Population Inferences and Variance Estimation for NAEP Data. *Journal of Educational Statistics, (17)*2, 175–190.

determining whether the sample differences are likely to represent actual differences among the groups in the population.

To determine whether a real difference exists between the average scale scores (or percentages of a certain attribute) for two groups in the population, one needs to obtain an estimate of the degree of uncertainty associated with the difference between the averages (or percentages) of these groups for the sample. This estimate of the degree of uncertainty, called the "standard error of the difference" between the groups, is obtained by taking the square of each group's standard error, summing the squared standard errors, and taking the square root of that sum.

$$SE_{A-B} = \sqrt{(SE_A^2 + SE_B^2)}$$

The standard error of the difference can be used, just like the standard error for an individual group average or percentage, to help determine whether differences among groups in the population are real. The difference between the averages or percentages of the two groups plus or minus 1.96 standard errors of the difference represents an approximately 95 percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between the groups in the population. If the interval does not contain zero, the difference between the groups is statistically significant at the .05 level.

The following example of comparing groups addresses the problem of determining whether the average mathematics scale score of group A is higher than that of group B. The sample estimates of the average scale scores and estimated standard errors are as follows:

Group	Average scale score	Standard error
A	218	0.9
B	216	1.1

The difference between the estimates of the average scale scores of groups A and B is 2 points (218 – 216). The standard error of this difference is

$$\sqrt{(0.9^2 + 1.1^2)} = 1.4$$

Thus, an approximately 95 percent confidence interval for this difference is plus or minus 1.96 standard errors of the difference:

$$2 \pm 1.96 \times 1.4$$
$$2 \pm 2.7$$
$$(-0.7, 4.7)$$

The value zero is within the confidence interval; therefore, there is insufficient evidence to conclude that group A's performance is statistically different from group B.

The procedure above is appropriate to use when it is reasonable to assume that the groups being compared have been independently sampled for the assessment.

Such an assumption is clearly warranted when comparing results for one state with another. This is the approach used for NAEP reports when comparisons involving independent groups are made. The assumption of independence is violated to some degree when comparing group results for the nation or a particular state (e.g., comparing national 2007 results for Black and White students), since these samples of students have been drawn from the same schools.

When the groups being compared do not share students (as is the case, for example, of comparing Black and White students), the impact of this violation of the independence assumption on the outcome of the statistical tests is assumed to be small, and NAEP, by convention, has, for computational convenience, routinely applied the procedures described above to those cases as well.

When making comparisons of results for groups that share a considerable proportion of students in common, it is not appropriate to ignore such dependencies. In such cases, NAEP has used procedures appropriate to comparing dependent groups. When the dependence in group results is due to the overlap in samples (e.g., when a subgroup is being compared to a total group), a simple modification of the usual standard error of the difference formula can be used. The formula for such cases is

$$SE^2_{Total-Subgroup} = \sqrt{(SE^2_{Total} + SE^2_{Subgroup} - 2pSE^2_{Subgroup})}$$

where p is the proportion of the total group contained in the subgroup. This formula was used for this report when a state was compared to the aggregate for the nation.

Conducting multiple tests

The procedures used to determine whether group differences in the long-term trend and main NAEP samples represent actual differences among the groups in the population and the certainty ascribed to intervals (e.g., a 95 percent confidence interval) are based on statistical theory that assumes that only one confidence interval or test of statistical significance is being performed. However, there are times when many different groups are being compared (i.e., multiple sets of confidence intervals are being analyzed).

For multiple comparisons, statistical theory indicates that the certainty associated with the entire set of comparisons is less than that attributable to each individual comparison from the set. To hold the significance level for the set of comparisons at a particular level (e.g., .05), the standard methods must be adjusted by multiple comparison procedures.[3] The procedure used by NAEP is the Benjamini-Hochberg False Discovery Rate (FDR) procedure.[4]

Unlike other multiple comparison procedures that control the family-wise error rate (i.e., the probability of making even one false rejection in the set of comparisons), the FDR procedure controls the expected proportion of falsely rejected hypotheses. Furthermore, the FDR procedure used in NAEP is considered appropriately less conservative than family-wise procedures for large families of compari-

sons.[5] Therefore, the FDR procedure is more suitable for multiple comparisons in NAEP than are other procedures.

Statistical comparisons of NAEP scores from different assessment years are made using a multiple comparison procedure. However, in figures 9, 11, 21, and 23, comparisons of the size of the Black-White achievement gap for each state to the national gap are made using pairwise comparisons, where each state is compared to the nation one at a time. For this reason, the results shown in these four figures may not correspond to results obtained from the NAEP Online Data Tool, which currently does not permit pairwise comparisons for this type of gap analysis.

Cautions in interpretation

It is possible to examine NAEP performance results for groups of students defined by various background factors measured by NAEP, such as race. However, a relationship that exists between achievement and another variable does not reveal its underlying cause, which may be influenced by a number of other variables. Similarly, the assessments do not reflect the influence of unmeasured variables. The results are most useful when they are considered in combination with other knowledge about the student population and the educational system, such as trends in instruction, changes in the school-age population, and societal demands and expectations.

[3] Miller, R.G. (1981). *Simultaneous Statistical Inference* (2nd ed.). New York: Spinger-Verlang.

[4] Benjamini, Y., and Hochberg, Y. (1995), op. cit.

[5] Williams, V.S.L., Jones, L.V., and Tukey, J.W. (1994, December) *Controlling Error in Multiple Comparisons with Special Attention to the National Assessment of Educational Progress.* Research Triangle Park, NC: National Institute of Statistical Sciences.

Appendix B: Supplemental Tables

Table B-1. Administration of NAEP national and state mathematics assessments, by grade: Various years, 1990–2007

Grade	1990 National	1990 State	1992 National	1992 State	1994 National	1994 State	1996 National	1996 State	1998 National	1998 State	2000 National	2000 State	2002 National	2002 State	2003 National	2003 State	2005 National	2005 State	2007 National	2007 State
4th grade	✔		✔	✔			✔	✔			✔	✔			✔	✔	✔	✔	✔	✔
8th grade	✔	✔	✔	✔			✔	✔			✔	✔			✔	✔	✔	✔	✔	✔

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, Various years, 1990–2007 Mathematics Assessments.

Table B-2.   Average national mathematics scale scores for all public school students at grades 4 and 8, by gender and eligibility for the National School Lunch Program: Various years, 1990–2007

	1990[n]	1992[n]	1996	2000	2003	2005	2007
All students							
Grade 4	212*	219*	222*	224*	234*	237*	239
Grade 8	262*	267*	269*	272*	276*	278*	280
Student Gender							
Grade 4							
Male	212*	220*	222*	225*	235*	238*	240
Female	211*	218*	222*	223*	233*	236*	238
Grade 8							
Male	262*	266*	270*	273*	277*	278*	281
Female	261*	267*	268*	271*	275*	277*	279
Student Eligibility for National School Lunch Program							
Grade 4							
Not eligible	—	—	‡	‡	244*	248*	249
Reduced-price lunch	—	—	‡	‡	230*	234*	236
Free lunch	—	—	‡	‡	220*	224*	226
Grade 8							
Not eligible	—	—	‡	‡	287*	288*	291
Reduced-price lunch	—	—	‡	‡	269*	270*	274
Free lunch	—	—	‡	‡	256*	260*	263

[n] Accommodations were not permitted for this assessment.
— Not available. Data were not collected prior to 1996.
‡ Reporting standards not met. Lunch eligibility data are not being reported in 1996 and 2000 because of the high percentage of students for whom information was not available.
* Significantly different (p<.05) from 2007.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Various years, 1990–2007 Mathematics Assessments.

Table B-3. Administration of NAEP national and state reading assessments, by grade: Various years, 1992–2007

Grade	1992 National	1992 State	1994 National	1994 State	1996 National	1996 State	1998 National	1998 State	2000 National	2000 State	2002 National	2002 State	2003 National	2003 State	2005 National	2005 State	2007 National	2007 State
4th grade	✔	✔	✔	✔			✔	✔	✔		✔	✔	✔	✔	✔	✔	✔	✔
8th grade	✔		✔				✔	✔			✔	✔	✔	✔	✔	✔	✔	✔

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, Various years, 1992–2007 Reading Assessments.

Table B-4. Average national reading scale scores for all public school students at grades 4 and 8, by gender and eligibility for the National School Lunch Program: Various years, 1992–2007

	1992[n]	1994[n]	1998	2000	2002	2003	2005	2007
All students								
Grade 4	215*	212*	213*	211*	217*	216*	217*	220
Grade 8	258*	257*	261	—	263*	261	260*	261
Student Gender								
Grade 4								
Male	211*	207*	210*	206*	214*	213*	214*	216
Female	219*	218*	215*	217*	220*	220*	220*	223
Grade 8								
Male	251*	250*	253*	—	258*	256	255*	256
Female	264	265	268*	—	267*	267	266	266
Student Eligibility for National School Lunch Program								
Grade 4								
Not eligible	—	—	‡	‡	‡	229*	230*	232
Reduced-price lunch	—	—	‡	‡	‡	211*	212*	215
Free lunch	—	—	‡	‡	‡	199*	201*	203
Grade 8								
Not eligible	—	—	‡	—	‡	271	270*	271
Reduced-price lunch	—	—	‡	—	‡	256	254	255
Free lunch	—	—	‡	—	‡	243*	245	246

[n] Accommodations were not permitted for this assessment.
— Not available. Data were not collected prior to 1996 or at grade 8 in 2000.
‡ Reporting standards not met. Lunch eligibility data are not being reported in 1998, 2000 and 2002 because of the high percentage of students for whom information was not available.
* Significantly different (p<.05) from 2007.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Various years, 1992–2007 Reading Assessments.

Acknowledgments

NATIONAL INDIAN EDUCATION STUDY

NIES 2011

The National Indian Education Study is designed to describe the condition of education for American Indian and Alaska Native students in the United States. This federally funded study is a collaborative effort to ensure that programs that serve American Indian and Alaska Native children are of the highest quality and meet their unique educational and culturally related needs.

The goal of the National Indian Education Study (NIES) is to collect information on academic achievement and educational experiences of American Indian and Alaska Native (AI/AN) students in order to understand and address their academic challenges. The results from the NIES are used in congressional testimony and serve as a benchmark for measuring the effectiveness of existing Native American programs. In addition, this ongoing study provides information on how AI/AN students progress in mathematics and reading over time.

The NIES, which is conducted in two parts, focuses on fourth- and eighth-grade students across the country. The study is designed to report results for the nation and for states that have relatively large populations of AI/AN students, and by school types and locations. The NIES provides information to help states, schools, and parents develop educational programs for AI/AN students, while respecting and honoring their distinct languages, cultures, and traditions. The educational programs ensure that AI/AN students meet the same challenging academic achievement standards as all other students across the country.

PART I

Part I of the study consists of the mathematics and reading portions of the National Assessment of Educational Progress (NAEP), which is administered to students nationwide. The results of NAEP are released as The Nation's Report Card, and are available for the nation, states, and in some cases, urban districts. NCES-contracted field staff coordinate with individual schools to schedule and administer the 90-minute assessment. As the federal government continues to develop programs to serve students in Native communities, results from the NIES, which is administered every two years, will assist in monitoring the progress of academic achievement.

PART II

Part II of the study is an Indian Education Survey, which asks AI/AN students, their teachers, and school principals about the inclusion of Native languages and cultural perspectives in the curriculum and about interactions between the school and the AI/AN community. The survey is administered immediately following the NAEP assessment. On average, it takes students up to 20 minutes to complete the survey; teachers and principals may need slightly longer. Data from the NIES 2011 questionnaires used in Part II will provide an understanding of the ways that cultural influences can affect the educational experiences of AI/AN students.

U.S. Department of Education
NCES 2010-469

:ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

BACKGROUND

The NIES is conducted as a part of NAEP, which was expanded to allow for more in-depth reporting on the achievement and experiences of AI/AN students. The NIES fulfills a mandate under Executive Order 13336 signed in 2004 to assist AI/AN students in meeting the challenging academic standards set forth in the Elementary and Secondary Education Act reauthorized in 2001 (Public Law 107-110) in a manner consistent with tribal languages, cultures and traditions. The NIES is conducted by the U.S. Department of Education's National Center for Education Statistics (NCES) within the Institute of Education Sciences, at the request of the U.S. Department of Education's Office of Indian Education (OIE) within the Office of Elementary and Secondary Education (OESE). The study is also supported by the Bureau of Indian Education, U.S. Department of the Interior.

NIES 2011

SCHEDULE

May - August 2010
Selected schools are asked to participate, a date is reserved for early 2011, and a school coordinator is appointed.

September - October 2010
School coordinators complete the fourth and eighth grade student information materials and arrange logistics for conducting the assessment.

January - March 2011
Experienced field staff visit the schools to administer the NAEP assessment (NIES Part I) and the survey (NIES Part II).

WHERE CAN I GET MORE INFORMATION?

NIES Information Line	1-888-747-4994
NIES website	http://nces.ed.gov/nationsreportcard/nies
National Assessment of Educational Progress (NAEP)	http://nces.ed.gov/nationsreportcard/
National Center for Education Statistics (NCES)	http://nces.ed.gov
Office of Indian Education (OIE)	http://www.ed.gov/about/offices/list/oese/oie/index.html

NIES 2011 Part I and Part II results will be available in Summer 2012
Results for both parts of the study can be found on the National Indian Education Study website and through the NAEP Data Explorer: http://nces.ed.gov/nationsreportcard/naepdata/.

:ies NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

NAEP

NATIONAL ASSESSMENT
OF EDUCATIONAL
PROGRESS

U.S. Department of Education
NCES 2014-487

Evaluación Nacional del Progreso Educativo

Visión General de NAEP

¿Qué es NAEP?

La Evaluación Nacional del Progreso Educativo (NAEP, por sus siglas en inglés) es la mayor herramienta de evaluación nacionalmente representativa y contínua de lo que los estudiantes en Estados Unidos saben y pueden hacer en diferentes materias. Desde 1969, NAEP ha proporcionado una medida común de los logros estudiantiles de todo el país. Los resultados se publican en La Libreta de Calificaciones de la Nación, y están disponibles para el país, los estados y en algunos casos, para distritos urbanos. NAEP se realiza por mandato del Congreso y lo administra el Centro Nacional para Estadísticas de la Educación (NCES, por sus siglas en inglés), dentro del Instituto de Ciencias de la Educación del Departamento de Educación de Estados Unidos.

¿En qué se diferencia NAEP de las evaluaciones de los estados?

NAEP cumple un papel diferente al de las evaluaciones de los estados. Cada estado tiene sus propias evaluaciones diseñadas para obtener información individual por estudiante acerca de sus logros en distintos estándares de contenido. NCES administra la misma evaluación NAEP en cada estado. Esto ofrece una medida común de los logros, lo que permite hacer comparaciones de los logros a nivel nacional y entre los estados y distritos urbanos participantes. NAEP no presenta los resultados de cada estudiante.

A diferencia de las evaluaciones estatales, también se les pide a los estudiantes, maestros y directores que participan en NAEP que contesten cuestionarios contextuales. Esta información proporciona un mejor entendimiento de las vivencias educativas de los estudiantes y los factores que podrían estar relacionados con el aprendizaje. Para averiguar más (en inglés) acerca de estos cuestionarios, vaya a http://nces.ed.gov/nationsreportcard/bgquest.aspx.

▸ El sitio Web de NAEP de NCES proporciona información (en inglés) más detallada acerca de la evaluación: http://nces.ed.gov/nationsreportcard

▸ Las herramientas de NAEP en Internet proporcionan acceso rápido y simple a datos de la evaluación NAEP, a preguntas de NAEP administradas anteriormente, a comparaciones de rendimiento y a más (en inglés): http://nces.ed.gov/nationsreportcard/about/naeptools.asp

▸ Puede encontrar copias completas (en inglés) de todos los informes en el sitio Web de La Libreta de Calificaciones de la Nación en: http://nationsreportcard.gov

¿Qué materias evalúa NAEP?

Las evaluaciones principales de NAEP se llevan a cabo en varias materias con estudiantes de cuarto, octavo y doceavo grado de todo el país. Las evaluaciones se realizan cada dos años en matemáticas y lectura, y cada cuatro años en ciencias y escritura. Otras materias, como artes, educación cívica, economía, geografía, conocimientos de tecnología e ingeniería, e historia de EE. UU., se evalúan periódicamente.

La evaluación NAEP de tendencias a largo plazo mide el rendimiento de los estudiantes en matemáticas y lectura, y se ha diseñado para garantizar comparabilidad a través de los años. Este tipo de evaluación de tendencias a largo plazo permite seguir el avance educativo desde principios de los 70. La evaluación se administra cada cuatro años a estudiantes de 9, 13 y 17 años de edad.

Los estudios especiales de NAEP, como el Estudio de calificaciones durante la secundaria (High School Transcript Study), se llevan a cabo periódicamente además de la evaluación principal y la de tendencias a largo plazo. La Junta Regidora de la Evaluación Nacional, la cual establece las normas de NAEP, determina el cronograma de las evaluaciones y el contenido para la medición. Para bajar una versión detallada (en inglés) en formato PDF del cronograma de evaluaciones, vaya a http://www.nagb. org/newsroom/assessment-schedule.htm.

¿Qué pueden esperar las escuelas y los estudiantes cuando participan en NAEP?

▶ NAEP se administra durante el horario escolar normal. A los estudiantes les toma de 90 a 120 minutos en completar la evaluación.

▶ A cada estudiante se le evalúa en una materia y se le pide que dé información contextual, como la cantidad de lectura que hace, los tipos de clases que toma y sus experiencias con la tecnología.

▶ Se acomodan los estudiantes con discapacidades y/o los estudiantes de inglés como segunda lengua o estudiantes con limitaciones lingüísticas en español.

▶ Los resultados de los estudiantes se mantienen de manera confidencial. Los representantes de NAEP son los encargados de traer a y llevarse de la escuela todos los materiales el día de la evaluación.

¿Cómo usa NAEP la tecnología para medir y dar a conocer las destrezas de una nueva generación de estudiantes?

A medida que las computadoras y herramientas digitales desempeñan un papel cada vez más importante en las aulas de hoy en día, NAEP sigue logrando avances con el uso de evaluaciones asistidas por tecnologías. El objetivo es dejar de usar las evaluaciones en papel para finales de esta década. A través de las siguientes evaluaciones innovadoras, NAEP está recolectando nuevos tipos de información que enriquecen nuestro entendimiento de lo que los estudiantes saben y pueden hacer, inclusive su uso de la tecnología para abordar la resolución de problemas.

▶ **EVALUACIÓN DE CONOCIMIENTOS DE TECNOLOGÍA E INGENIERÍA (TEL, POR SUS SIGLAS EN INGLÉS):** TEL marca una nueva frontera en NAEP y en evaluaciones en gran escala. Es una evaluación por computadora e interdisciplinaria que desafía a los estudiantes mediante la realización de actividades interactivas y de resolución de problemas basadas en situaciones que reflejan la realidad. La evaluación TEL mide qué tan bien los estudiantes entienden y ponen en práctica los principios de tecnología e ingeniería en situaciones de la vida real. Para averiguar más (en inglés) sobre TEL, vaya a http://nces.ed.gov/nationsreportcard/tel.

▶ **EVALUACIÓN DE ESCRITURA:** La evaluación de escritura se administra por computadora y les pide a los estudiantes que contesten pautas de escritura presentadas en formatos multimedia que incluyen audio y videos cortos. Además de proporcionar puntuaciones de escritura, los resultados de las evaluaciones proporcionan información acerca de la medida en que los estudiantes realizaron ciertas acciones en la computadora al responder estas tareas, tal como el uso del diccionario de sinónimos. Los resultados y la información contextual están disponibles (en inglés) en http://nces.ed.gov/nationsreportcard/writing.

▶ **ACTIVIDADES INTERACTIVAS EN COMPUTADORA DE CIENCIAS (ICT, POR SUS SIGLAS EN INGLÉS):** La evaluación de ciencias de NAEP incluye actividades ICT que desafían a los estudiantes a resolver problemas científicos y hacer experimentos, frecuentemente con técnicas de simulación. En comparación con la evaluación de lápiz y papel, la evaluación con actividades ICT proporciona al estudiante mayores oportunidades para demostrar destrezas que se emplean en las ciencias sin muchas de las limitaciones logísticas asociadas con un ambiente de laboratorio o natural. La biblioteca completa de actividades ICT (en inglés) publicadas que se utilizaron en la evaluación de 2009 está disponible en http://nationsreportcard.gov/science_2009/ict_tasks.asp.

¿Por qué es valiosa la participación estudiantil?

La participación de todos los estudiantes seleccionados permite que NAEP proporcione una descripción precisa del rendimiento estudiantil en Estados Unidos. Ya que NAEP no se diseña para presentar los resultados de cada estudiante o escuela, no es necesario que cada estudiante de cada escuela tome la evaluación. En cambio, NCES utiliza una muestra aleatoria para asegurarse de que los participantes de NAEP representen la diversidad geográfica, racial/étnica y socioeconómica de las escuelas y de los estudiantes a nivel nacional. Cada participante representa a cientos de otros estudiantes. La participación de cada uno de ellos es vital para recolectar y compartir información válida acerca de los logros de los estudiantes. Los maestros, directores, padres, legisladores y los investigadores utilizan los resultados de NAEP para evaluar el progreso de los estudiantes y desarrollar maneras para mejorar la educación a nivel nacional.

Después de cada evaluación, algunas preguntas de NAEP se dan a conocer al público en general junto con información acerca del rendimiento de los estudiantes para cada pregunta. A continuación se muestran algunos ejemplos de preguntas de NAEP y respuestas de los estudiantes.

Cuarto Grado, Matemáticas

En un frasco hay 6 cubitos del mismo tamaño.
2 cubitos son amarillos
3 cubitos son rojos
1 cubito es azul

Carlos va a sacar un cubito sin mirar.
¿De qué color es más probable que sea el cubito que saque?

rojo

¿Cuál es la probabilidad de que este color sea sacado?

3 de 6

Fourth Grade, Mathematics

There are 6 cubes of the same size in a jar.
2 cubes are yellow.
3 cubes are red.
1 cube is blue.

Chuck is going to pick one cube without looking. Which color is he most likely to pick?

red

What is the probability of this color being picked?

3 out of 6

Encuéntrenos en:

Obtenga información sobre NAEP en todo momento con la aplicación *NAEP Results* (resultados de NAEP) para su celular.

www.ed.gov   ies.ed.gov

THE NATIONAL CENTER FOR EDUCATION STATISTICS

FOCUS on NAEP

How Does NAEP Ensure Consistency in Scoring?

Abstract: Each NAEP assessment requires the scoring of thousands, and often millions, of written responses to open-ended questions. NCES and its contractors have devised a variety of techniques to ensure that these heterogeneous responses are scored consistently.

The National Center for Educational Statistics (NCES) has been conducting the National Assessment of Educational Progress (NAEP) since 1969. In addition to regular assessments in reading, mathematics, science, and writing, NCES also conducts assessments in such subjects as geography, U.S. history, civics, and the arts.

All of these assessments include constructed-response questions in addition to multiple-choice items. Many include "short constructed-response" questions, which require students to provide a numerical response or write a few words or sentences, as well as "extended constructed-response" questions, which may require the student to write a paragraph or more, perform a science experiment and write a description of what was done, or solve a word problem in mathematics, providing a written explanation of the answer. Writing assessments require students to produce two extensive writing samples, while the arts assessments require students to create and perform art.

Extended constructed-response questions for NAEP assessments such as reading, U.S. history, geography, and civics are scored according to four-level scoring guides. Four-point answers are typically scored as incorrect, partial, essential, and fully correct. However, some assessments, such as the arts, mathematics and writing assessments, have questions that recognize five or even six levels of performance.

Each national assessment generates thousands of student responses that must be scored individually, and combined state/national assessments can generate almost five million responses.[1] NCES and its contractors have developed a large number of special techniques to ensure that these constructed-response questions can be scored consistently. This *Focus on NAEP* will discuss the techniques used to score written assessments such as reading, mathematics, writing, and science. A separate *Focus on NAEP* will cover the special problems encountered in assessing the arts.

Selecting Scorers

In the year 2000, NCES will conduct two national/state assessments, in mathematics and science, at grades 4, 8, and 12 at the national level and at grades 4 and 8 at the state level. In addition, there will be a national reading assessment for grade 4 only. The three assessments will generate close to 10 million constructed responses. The scoring will be done, as it has been done for previous assessments, by National Computer Systems (NCS). Educational Testing Service (ETS) develops the scoring guides for the questions and provides training in their use.

Scoring will be done at two on-line Professional Scoring Centers, one in Iowa City and the other in Tucson, Arizona. The contractors will hire about 150 scorers for the mathematics assessment, about 175 for the science, and about 50 for the reading.

Scorers selected for the assessment will have the following qualifications:

- A minimum of a bachelor's degree in the appropriate academic discipline (mathematics, science, or English), or in education;

- Scoring experience in NAEP or non-NAEP assessments preferred;
- Teaching experience at the elementary or secondary level preferred.

The 2000 Mathematics Assessment will have bilingual booklets for the 4th and 8th grades. Scorers fluent in Spanish will be hired for the scoring of booklets answered in that language.

Training Scorers

Training scorers to score short and extended constructed-response questions consistently is one of the most important parts of the entire scoring procedure. There is separate training for each constructed-response question.[2]

Training involves the following:

- Presenting and discussing the question to be scored and the question's rationale;
- Explaining the scoring guide to the team and discussing the "Anchor Packet," which contains the scoring guide, the question, its scoring rationale, and the "Anchor Set" of student responses that represent the various score points in the guide;
- Discussing the rationale behind the guide, focusing on the criteria that differentiate the levels in the guide;
- Practicing scoring on a "Practice Set" of students' answers;
- Continuing to practice until a consensus is reached on how to apply the scoring guide.

Trainers and participating experts in the field begin by selecting from 150 to 300 student answers to an extended constructed-response question. They score them all, for training purposes, and use the answers to create three different training sets, the Anchor Set, the Practice Set, and the Qualification Set.

Answers in the Anchor Set have the scores written on them. An Anchor Set contains at least three answers for every score point in a question. The Anchor Set for a three-point question will usually have 10 answers, and the Anchor Set for a four-point question will have about 15. The trainers also score a Practice Set of about 10 to 20 answers, and a Qualification Set of similar size, but do not put the scores on the answers.

Scorers, divided into training teams, will first study the scoring guide developed for a given question. Then they receive the Anchor Set of answers, which they review in conjunction with the scoring guide. Then they are given the Practice Set. Scorers score each of the answers, and then are given the "true" score, arrived at earlier by the trainers, for comparison and discussion.

Once the scorers are familiar with the scoring of a question, they are given a Qualification Set of answers to score. At least 80 percent of their scores must match the scores given by the trainers. Scorers who fail to get 80 percent discuss the scoring of the Qualification Set with their trainer and then are given a second Qualification Set. If they fail to get at least an 80 percent match on this set, they cannot score the question.

Image Scoring and Monitoring

Scoring of constructed-response questions is done by an "Image" process. While student answers are written in traditional answer booklets, for scoring purposes they are converted into computer images. This allows all the answers for a given question to be grouped together and scored at the same time. Scorers are trained to score the answers to a question, and then work exclusively on answers to that question until each one has been scored.

When scorers begin scoring answers to a question, they first take turns scoring the same question, comparing answers, or score in pairs as a final quality check before scoring on their own. They receive retraining at the beginning of each day and after any break that exceeds 15 minutes.

Scorers will be monitored by supervisors (known as "table leaders") in a variety of ways. A certain percentage of answers for constructed-response questions will be scored twice.[3] The second scorer will not know the score assigned by the first scorer. Because all scoring is done on a linked computer network, table leaders will have data on the scoring agreement rates for all scorers while the scoring is in progress. Figure 1 provides a "reliability summary" used to keep track of scoring consistency.

A minimum standard agreement rate will be set for each question, which will take into account both the number of score points for a question and the subject being assessed. For example, a higher agreement rate is set for a three-point question than a four-point question; and agreement rates will be higher for a subject such as mathematics, where the "correct" answer can usually be defined with greater precision, than for a subject such as reading. In 1998, the average standard agreement rate for questions on the reading assessment was 91 percent for grade 4, 90 percent for grade 8, and 89 percent for grade 12. For the 1996 mathematics assessment, it was 96 percent for all three grades.

If the minimum agreement rate is not met for a question, a number of different remedial actions may be necessary. If all or most members of a scoring team appear to be below the average, retraining may be appropriate. If there seems to be a problem with one scorer, the scorer may be reassigned.

The answers that were scored with insufficient agreement rates need to be rescored. This may be done by a group of supervisors, or all the scores for a question may be erased, and the team starts over again. Sometimes, the question is assigned to a different scoring team.

Occasionally, the scoring trainer may decide that the scoring guide needs to be refined, although this rarely happens during an assessment. Scoring guides are more likely to be refined during preliminary testing of assessment questions.

Table leaders will have methods to review an individual scorer's consistency as well as the consistency of a scoring team. A table leader will typically review 10 percent of the answers scored by a scorer, and will discuss with the scorer any score that appears inappropriate. A table leader has the authority to rescore any answer, although this does not affect the inter-rater reliability data. To check on scoring consistency across individual scorers, a table leader can also review all the answers that were given a particular score by a scoring team or the committee that developed the assessment questions.

The NAEP assessments that NCES will be conducting in 2000 are periodically redesigned to keep them responsive to changes in curricula and also to reflect improvements in assessment techniques. However, because NCES uses the same assessment instrument several times before making changes, these assessments usually offer some

Figure 1.—Reader Reliability Summary

First Scorer	Blank		1		2		3		4		Illegible		Off Task	
Second Scorer	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Blank	115	100%												
1			18	90%	2	10%								
2			2	1%	330	95%	17	5%						
3					7	4%	156	92%	6	4%				
4							3	13%	21	88%				
Illegible														
Off Task													1	100%

Total Times 2nd Read: 678   Percent Agreement: 94.5%

This sample "Reader Reliability Summary" shows how table leaders at National Computer Systems keep track of the scoring consistency of the second scoring of a single NAEP extended constructed-response question.

The sample summary is for a four-point question, whose answers are scored as either "incorrect," "partial," "essential," or "fully correct"—"fully correct" answers receiving the full four points. (The rows and columns marked "Blank," "Illegible," and "Off Task" are for answers that are unscorable due to omission, completely illegible handwriting, and unresponsiveness to task.)

This summary shows the cumulative agreement rate for all second scoring of students' answers to a single four-point question. Scoring decisions by the first scorer head the double columns at the top of the chart, while those for the second scorer, appearing in the far left-hand column, govern the rows. The chart should be read row by row. (The "3" row has been bolded for illustration.)

The cells created by the intersection of the "3" row and the double columns labeled "2", "3," and "4"give information on answers that received a "3" score from the second scorer. The first "n" or "number" cell shows that 7 answers scored as "3" by the second scorer received a score of "2" from the first scorer. The first "%" cell indicates that these 7 answers constitute 4% of the answers scored as "3" by the second scorer.

The next two cells to the right indicate that 156 answers, or 92% of all the answers receiving a "3" score from the second scorer, received a "3" from the first scorer as well. The next two cells indicate that 6 answers (4%) received a "3" from the second scorer and a "4" from the first scorer.

Ideally, all numbers and percentages would be in the shaded cells, and all percentages would be 100%. In fact, however, this only occurs for the "Blank" and "Off Task" answers. The "Percent Agreement" of 94.5% seen in the lower right-hand corner is obtained by dividing the total number of "agreed" scores (641) by the total number of scores (678).

3

trend data. For this reason, decisions by scorers working on the current assessments will be compared with decisions by past scorers when appropriate. A similar procedure is used for the NAEP long-term trend assessments, whose primary function is to track student performance over time.

Conclusion

Achieving consistency in the scoring of constructed-response questions begins with the selection of individuals who have a background in education and experience in scoring. These individuals are trained carefully in the scoring of each question, so that all the scorers, working independently, give the same number of points to any answer to that question. Regular second scoring of answers to every question ensures that this consistency is maintained throughout the scoring process.

Endnotes

[1] The NAEP 1997 arts assessment (in music, theatre, and the visual arts) covered the 8th grade only, and involved a total of about 6,500 students. The arts assessment involved relatively few questions, because students devoted much of their time to a single creating or performing task. A national/state assessment in a subject such as science will involve about 7,500 students at each of three grades (4th, 8th, and 12th), plus about 2,500 per state per grade. In the past, more than 40 states and other jurisdictions have participated in each NAEP state assessment.

[2] The training procedures described are for extended constructed-response questions. The procedures for short constructed-response questions are similar but less elaborate.

[3] Six percent of the answers for the constructed-response questions of the mathematics and science assessments for grades 4 and 8 will be scored twice. This will include both the national and state assessments for these subjects and grades. In addition, 25 percent of the answers for the grade 12 assessments in science and mathematics will

be scored twice, a procedure that will also be followed for the reading assessment (grade 4 only). A larger percentage will be scored for these assessments because they are national assessments only, and thus will involve substantially fewer answers.

For Further Information

The NAEP 1996 Technical Report, NCES 1999–452, discusses all technical aspects of the 1996 Mathematics and Science Assessments and the 1996 Long-Term Trend Assessments.

Technical Report: NAEP 1996 State Assessment Program in Science, NCES 1998–480, covers the technical aspects of that assessment in detail.

Single copies of both reports are available free from ED Pubs, P.O. Box 1398, Jessup, Md. 20794–1398. Copies may also be downloaded from the World Wide Web *http://nces.ed.gov/pubsearch/*

The *Focus on NAEP* series briefly summarizes information about the ongoing development and implementation of the National Assessment of Educational Progress (NAEP). The series is a product of the National Center for Education Statistics (NCES), Gary Phillips, Acting Commissioner, and Peggy Carr, Associate Commissioner for Education Assessment. This *Focus on NAEP* issue was written by **Sheida White** of NCES, **Connie Smith** of National Computer Systems, and **Alan Vanneman** of the Education Statistics Services Institute.

To order other NAEP publications, call toll free 1–877–4ED–Pubs (1–877–433–7827), TTY/TDD 1–877–576–7734;
E–mail: edpubs@inet.ed.gov;
Internet: *http://www.ed.gov/pubs/edpubs.html*

The NCES World Wide Web Home Page is: *http://nces.ed.gov/*

:ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

NAEP
NATIONAL ASSESSMENT
OF EDUCATIONAL
PROGRESS

U.S. Department of Education
NCES 2010-470

NAEP Writing Computer-Based Assessment

An Overview for Grades 8 and 12

What to Expect for the NAEP Writing Computer-Based Assessment

What should **schools** expect?

As with other NAEP assessments, NAEP staff will administer the assessment and work with school staff to organize the writing assessment activities. NAEP staff will bring necessary materials, including laptop computers and earbuds, to the school on assessment day. Schools will only need to provide one room, desks or tables, and electricity; schools do not have to provide Internet access.

About 28-30 students in each school will be selected to participate. The assessment will be administered in two sequential sessions of about 15 students. No other NAEP subject area assessments will be given in schools participating in the writing computer-based assessment.

What can **students** expect?

Before the assessment begins, students will be shown a tutorial which will help them become familiar with the way material is presented on the computer screen and how to use the program. Students will compose their responses in a word processing program similar to the programs they frequently use. They will be able to use common tools for editing, formatting, and text analysis; they will not have access to irrelevant or distracting tools such as clip art, font type and color, or the Internet. Students may ask questions anytime before and during the assessment.

It will take approximately 90 to 120 minutes for students to complete the assessment, 60 minutes of which is allotted to writing two responses. Students will respond to text, video or animated prompts on the computer and listen to audio prompts via earbuds, which will be provided. The assessment questions are designed to measure students' ability to persuade, to explain, or to convey experience.

Students will also respond to questions designed to gather information that provides context for understanding the assessment results, such as how many pages the student has to read each day for school and homework, and whether there is a computer in their home that they can use. New to the assessment are specific questions about computer use, such as writing for school assignments, writing e-mails, and using the Internet.

In today's society, writing with paper and pencil has largely been replaced by writing using a computer. Students are expected to compose on a computer as they move through school and into the workforce. Reflecting the changes in technology, eighth- and twelfth-grade students taking the **National Assessment of Educational Progress (NAEP) writing assessment** will use a computer to compose and edit their responses.

→ How is the NAEP writing computer-based assessment different from past writing assessments?

The NAEP writing assessment has always been paper-and-pencil based for grades 4, 8, and 12. In 2011, for the first time, the NAEP writing assessment will measure students' ability to write using a computer at grades 8 and 12. The assessment is computer-based, and is designed to take advantage of many features of current digital technology, such as word processing software. The computer-based writing tasks are delivered in multimedia formats, such as short videos and audio.

93% of eighth-graders who took the NAEP reading assessment in 2009 reported that they used a computer at home.

(National Assessment of Educational Progress, 2009.)

→ Why is the writing assessment computer-based for grades 8 and 12?

The 2011 NAEP Writing Framework mandated that the writing assessment be designed to reflect the way today's students compose and are expected to compose: using a computer. For example, in 2007, NAEP data showed that 75% of twelfth-graders who took the writing assessment reported that they almost always used a computer to make changes to a paper or report (for example, spell-check, cut and paste). The writing assessment was developed to take into account the new role that technology plays as students move through school and into the workforce.

→ How was the new writing framework developed?

The 2011 NAEP Writing Framework was developed by the National Assessment Governing Board, with ideas from a wide range of individuals and organizations, including writing experts, school administrators, policymakers, teachers, parents, and others. To view the framework, visit http://www.nagb.org/publications/frameworks.htm.

→ Will accommodations be provided?

NAEP assessments strive to include as many students as possible. The writing computer-based assessment system is designed to comply with federal policy, which instructs that electronic information technology be accessible to all people, including those individuals with disabilities. The computer program for the writing assessment features optional settings that will reduce the need for accommodations, among them:

- A 'Speak' function on computers that can read aloud all or selected portions of directions, tasks, and stimulus materials. (Some stimulus materials will not respond to a 'Speak' function but will have separate play buttons.)
- Adjustable font size up to 48 point (letters about 2/3 inch tall).
- Choice of color schemes, including a high contrast option.
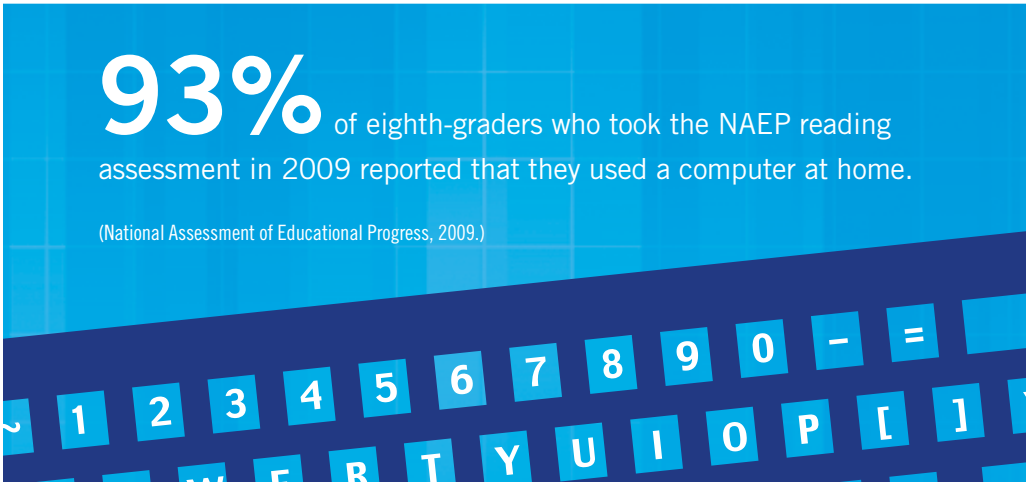
As with other NAEP assessments, accommodations such as alterations in the testing setting or timing will be provided. Any change that alters the skill NAEP is designed to measure will not be allowed. If needed, for the computer-based writing assessment, accommodations will be available, such as screen magnification greater than 48 point font, and the ability to enlarge graphics. Accommodations such as bilingual word-to-word dictionaries, or signed directions, can be provided by the school staff.

➔ How will the responses be scored?

Trained scorers will evaluate the responses. In NAEP, scorers are taught to use the scoring rubric through extensive training using many example responses. They will evaluate three broad features of writing, as follows:

- Development of Ideas
- Organization of Ideas
- Language Facility and Use of Conventions

Responses will be evaluated holistically; individual elements or parts of the response will not be scored separately.

The National Assessment of Educational Progress (NAEP) is a continuing and nationally representative assessment of what our nation's students know and can do in core subjects such as mathematics, reading, science, and writing. NAEP was first administered in 1969 to measure student achievement nationally. Teachers, principals, parents, policymakers, and researchers all use NAEP results to assess progress and develop ways to improve education throughout the country. The results of NAEP are released as The Nation's Report Card, and are available for the nation, states, and in some cases, urban districts. For more information, please visit http://nces.ed.gov/nationsreportcard/.

For more information about NAEP, visit http://nces.ed.gov/nationsreportcard.

www.ed.gov   ies.ed.gov

An Overview of Procedures for the NAEP Assessment

U.S. Department of Education
Institute of Education Sciences
NCES 2009-493

NAEP • **The National Assessment of Educational Progress**

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
Sue Betka
Acting Director

National Center for Education Statistics
Stuart Kerachsky
Acting Commissioner

table of contents

Mandated by Congress, the National Assessment of Educational Progress (NAEP) surveys the educational accomplishments of U.S. students and monitors changes in those accomplishments. NAEP tracks the educational achievement of fourth-, eighth-, and twelfth-grade students over time in selected content areas. Since 1969, NAEP has been collecting data to provide educators and policymakers with accurate and useful information. NAEP gives a comprehensive picture of how students are doing year after year. It has become widely known as "The Nation's Report Card."

About NAEP

The National Assessment Governing Board sets policy for NAEP, and the Commissioner of Education Statistics, who heads the National Center for Education Statistics (NCES) in the U.S. Department of Education's Institute of Education Sciences, is responsible for carrying out the assessment. Within NCES, the Associate Commissioner for Assessment executes the program operations and ensures technical quality control. Under the direction of the Associate Commissioner, contractors carry out the development, administration, scoring, and analysis of NAEP.

Over a million students are assessed to provide achievement data for fourth- and eighth-graders representative of all states, the District of Columbia, Department of Defense schools, and selected urban districts. The assessment is administered by NAEP contract employees and testing and admintrative procedures together require about 90 minutes of each student's time. School administrators and teachers also fill out questionnaires as part of the assessment.

NAEP has produced hundreds of reports in its history, chronicling trends over time in the performance of 9-, 13-, and 17-year-olds and fourth-, eighth-, and twelfth-grade students. NAEP also releases state-level results for certain assessments and district-level results for some jurisdictions. NCES strives to present findings in the most accurate and useful manner possible, publishing reports designed for the general public and specific audiences and making the data available to researchers for secondary analyses.

NAEP reports do not advocate specific pedagogies or policies. Instead, NAEP reports describe student performance in the context of the educational system in ways that inform discussion among policymakers and educational leaders. NAEP is not intended to drive state or local standards, tests, and curricula. By law, the federal government may not use NAEP to establish, require, or influence state or local educational standards, assessments, curriculum, classroom materials, or instructional practices. States or districts, however, may voluntarily draw from NAEP frameworks, assessments, or procedures when contemplating changes in their own programs.

Comprehensive information about NAEP, including assessment results, background questionnaires, and sample questions, can be found on the web at http://nationsreportcard.gov or http://nces.ed.gov/nationsreportcard. Subject framework information and additional, NAEP-related materials can be found at the Governing Board's website (http://www.nagb.org).

introduction

introduction

About This Guide

The goals of this publication are to provide readers with an overview of the project and to help them better understand the philosophical approach, procedures, analyses, and psychometric underpinnings of NAEP.

The guide follows a question-and-answer format, presenting the most commonly asked questions and following them with succinct answers. A glossary is found at the end of this guide; users can reference this glossary for more information on bold-faced words.

Q: *What is NAEP?*

A: Often called "The Nation's Report Card," the National Assessment of Educational Progress (NAEP) is the only nationally representative, continuing assessment of what America's students know and can do in various subject areas. As a congressionally mandated project of the National Center for Education Statistics (NCES) within the U.S. Department of Education and the Institute of Education Sciences (IES), NAEP provides a comprehensive measure of students' learning at critical junctures in their school experiences.

Since 1969, NAEP has conducted regular assessments and made objective information about student performance available to both policymakers and the general public, thereby playing an integral role in evaluating the condition and progress of the nation's educational outputs. NAEP is a voluntary assessment that collects only information related to academic achievement. NAEP is required by law to guarantee the confidentiality of all data related to individual participating students and their families. Results are reported based on the average performance of students at the national or state level. The NAEP assessments are not designed to permit the reporting of information regarding individual students or schools.

Further Details

Overview of NAEP

Since 1969, NAEP assessments have been conducted periodically in such subjects as reading, mathematics, science, writing, U.S. history, civics, economics, geography, and the arts.

As head of NCES in the U.S. Department of Education, the Commissioner of Education Statistics is responsible by law for carrying out the NAEP program. The National Assessment Governing Board establishes policy for the program. Although its members are appointed by the Secretary of Education, the Governing Board is independent of the department.

NAEP does not provide scores for individual students or schools; instead, it offers results regarding subject-matter achievement, instructional experiences, and school environment for populations of students (e.g., fourth-graders) and **student groups** of those populations (e.g., female students, Hispanic students). NAEP results are based on a representative **sample** of student populations of interest defined by, for example, grade level, race/ethnicity, or gender.

Between 1969 and 1979, NAEP conducted at least one assessment every year. From 1980 to 1996, assessments were administered once every 2 years. In 1996, NAEP returned to annual assessments. In 1990, Congress authorized NAEP to initiate state-level assessments, enabling states that chose to participate to compare their results with those of the nation and other participating states. The *No Child Left Behind Act of 2001* placed greater emphasis on state NAEP by mandating states to participate in biennial mathematics

THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP)

MAIN

LONG-TERM TREND

NATIONAL
Sample:
Public and Non-public
Grades 4, 8 and 12

Subjects:
Reading,
mathematics, science,
writing, U.S. history,
economics, civics,
geography, the arts…

(See page 55 for a
complete list of current and
planned subjects.)

STATE
Sample:
Public
Grades 4 and 8

Subjects:
Reading,
mathematics,
science, and
writing

TRIAL URBAN DISTRICTS
Sample:
Public
Grades 4 and 8

Subjects:
Reading,
mathematics,
writing and
science

NATIONAL
Sample:
Public and
Non-public
9-, 13- and 17-
year olds

Subjects:
Reading and
mathematics

and reading assessments in grades 4 and 8 as a condition for receiving **Title I** funds. (Title I of the Elementary and Secondary Education Act provides federal assistance to eligible schools and districts to help children who are at risk of not meeting education standards.)

The NAEP program includes two distinct components: "main NAEP" and "long-term trend NAEP." Main NAEP includes assessment instruments that have typically been developed since the early 1990s and are used at both the national and state levels. Long-term trend NAEP includes assessment instruments that date back to as early as 1969. Long-term trend NAEP is administered at the national level only and is administered less frequently than main NAEP. The above figure displays the various components of the NAEP program.

Even though main NAEP and long-term trend NAEP both assess reading and mathematics, these two program components use distinct data collection procedures, separate samples of students defined by different criteria, and different test instruments based on different **frameworks**. The **background questionnaires** that are used to collect information about students' instructional experiences and their school environments also vary between the main and long-term trend assessments. The results from these two assessments are reported separately, and are not directly comparable.

Main NAEP (national and state)

The term "main NAEP" is used to refer to national and state levels of the program that utilize the same assessment instruments based on the most recently developed frameworks. For the nation, results are reported for students from both public and non-public schools and for specific census-defined geographic **regions** of the country (Northeast, South, Midwest,

and West), as well as for several major demographic student groups. At the state and district levels, results are currently reported for public school students only and are broken down by the same demographic student groups as used for reporting national results.

The main NAEP assessments follow assessment frameworks developed by the Governing Board and use the latest advances in assessment methodology. Indeed, NAEP has pioneered many of these advances. The assessment instruments are flexible, so they can be adapted to changes in curricular and educational approaches. For example, main NAEP assessments include **constructed-response question**s (questions that ask students to write responses ranging from a single word or figure to a few paragraphs) and questions that require the use of calculators and other materials.

Recent main NAEP assessment instruments have typically been kept stable since the early 1990s, allowing short-term trend results to be reported. For example, the 2003 fourth-grade reading assessment has followed a short-term trend line that began in 1992 and continued in 1994, 1998, 2000, 2002, 2003, 2005, and 2007. However, the Governing Board has revised and updated the reading framework for use in the 2009 assessment, which will mark the start of a new trend line. Frameworks for other subjects are typically updated every ten years, such as math, which was revised in 2005.

The main assessments report results for grade samples–grades 4, 8, and 12 at the national level and grades 4 and 8 for state and participating

urban districts. They periodically measure students' achievement in a variety of **subject areas**. Reading, mathematics, science, and writing are assessed with samples representative of the nation and participating states. Other subject areas, such as U.S. history, civics, economics, and geography, are assessed only at the national level. (See page 55 for a list of subjects assessed by NAEP and the schedule of assessments.)

Initially, NAEP was a national-level assessment only. The national samples were not designed to support the reporting of accurate and representative state-level results. In 1988, however, Congress passed legislation authorizing a voluntary Trial State Assessment (TSA). In 1996, "Trial" was dropped from the title of the state assessments based on congressionally mandated evaluations. The first TSA occurred in 1990, and approximately 90 percent of states participated. The District of Columbia, the Department of Defense Education school system, Puerto Rico, and the Bureau of Indian Education schools joined the assessment in subsequent years.

The *No Child Left Behind Act of 2001* strongly encourages states to participate in biennial fourth- and eighth-grade NAEP reading and mathematics assessments beginning in the 2002–2003 school year in order to provide the residents of each state with reliable and valid information regarding the academic progress of their students. Both subjects are tested in the same year. Under the legislation, all states and school districts must agree to participate in these assessments in order to receive full funding from the federal Title I program. The law relieves states of NAEP's financial and administrative burden by providing federal funds to pay all costs involved in coordinating and administering the NAEP assessments.

question 1

Federal appropriations authorized for the No Child Left Behind Act supported the development of the Trial Urban District Assessment (TUDA) in 2002. TUDA is designed to explore the feasibility of using NAEP to report on the performances of fourth- and eighth-grade public school students at the district level. The number of districts assessed has grown from five, in 2002, to eleven, in 2007, and the original subjects—reading and writing—have been supplemented by assessments in mathematics and science. Participating students take the same tests as those participating in the main NAEP assessment, and they constitute a representative sample of their districts. The results from TUDA make it possible to compare the performance of students in participating urban school districts to that of public school students in the nation, in large central cities, and to each other.

For further information about state and jurisdiction participation in state NAEP and subjects assessed, consult the NAEP website (http://nces.ed.gov/nationsreportcard).

Long-Term Trend NAEP

The long-term trend assessments report results for age samples (9-year-olds, 13-year-olds, and 17-year-olds). In the past, these assessments have measured students' achievements in mathematics, science, reading, and writing. Currently, only mathematics and reading are assessed for long-term trend NAEP.

Measuring trends in student achievement, or change over time, requires that past procedures be replicated as precisely as possible. Therefore, instruments for the long-term trend assessment developed in the 1960s, 1970s, and 1980s were maintained until 1999 to provide consistent measurement over long periods of time. Up until 1999, the long-term trend assessment administered these instruments every few years. For the 2004 long-term trend assessment, however, it was decided that assessment instruments and procedures should be made consistent with the designs and procedures used in the main NAEP assessment. In order to ensure that assessment results could be interpreted consistently over time, a bridge study was conducted. A bridge study involves comparing two assessments: one that replicates the assessment given in the previous assessment year (a bridge assessment) and one that represents the new design (a modified assessment). In 2003–2004, students were randomly assigned to take either the bridge or modified assessment. The bridge assessment replicated the instrument given in 1999 and used the same administration procedures. The modified assessment included new items and features modeled after the main NAEP assessment. The modified assessment provides the basis of comparison for all future assessments, and the bridge links its results to the results of the previous 30 years.

Background Questionnaires

While the primary focus of NAEP is on achievement in specific subject areas, NAEP collects a wealth of other information to address many questions about student achievement. How well prepared are today's teachers? How much homework are students assigned? How do schools vary in terms of courses offered? NAEP attempts to address these questions and others through data collected on background questionnaires.

Sampled students, as well as their teachers and principals, complete these questionnaires to provide NAEP with data about students' school backgrounds and educational activities. Students answer questions about courses, homework, and a limited number of additional factors related to instruction. Teachers answer questions about their professional qualifications and teaching activities, while principals answer questions about school-level practices and policies. Relating student performance on the subject-related portions of the assessments to the information gathered on the background questionnaires increases the usefulness of NAEP findings and provides a context for understanding student achievement.

Related Questions:

Question 2: *What subjects does NAEP assess? How are the assessment questions determined?*

Question 5: *How many schools and students participate in NAEP, and who are they? When are the data collected during the school year?*

question 1

Q: *What subjects does NAEP assess? How are the assessment questions determined?*

A: Since its inception in 1969, the National Assessment of Educational Progress (NAEP) has assessed numerous academic subjects, including mathematics, reading, science, writing, geography, U.S. history, economics, civics, and the arts. (A chronological list of assessments planned through 2017 appears on page 55.)

The National Assessment Governing Board selects the subjects to be assessed and oversees creation of both the frameworks that underlie the NAEP assessments and the specifications that guide the development of the assessment instruments. The framework for each subject area is determined through a collaborative process involving teachers, curriculum specialists, subject-matter specialists, school administrators, parents, and members of the general public. The specifications provided by the Governing Board bridge the gap between the frameworks and the assessments by indicating the way in which the intent of the framework is to be implemented during item development.

Further Details

Selection of Subjects

In 1988, the legislation authorizing NAEP charged the Governing Board with determining which **subject areas** to assess and setting the schedule for the assessments. Beginning with the 2003 assessment, state NAEP included biennial mathematics and reading assessments for grades 4 and 8. Since 2002, TUDA has assessed urban districts in reading, writing, mathematics, and science. Other subjects NAEP has assessed include civics, U.S. history, economics, geography, and the arts. The table on page 55 lists NAEP's planned assessments through 2017.

Development of Frameworks

Frameworks are the blueprints that the Governing Board uses to specify the content and guide the development of assessment instruments in each subject. The validity of educational inferences made using NAEP data is dependent on the implementation of high standards and rigorous procedures for framework development.

Developing a particular framework involves the following elements:

- widespread participation and reviews by educators and state education officials in the field of interest;

- reviews by steering committees whose members represent policymakers, practitioners, and members of the general public;

- involvement of subject supervisors from the **education agencies** of prospective participants;

- public hearings; and

- reviews by National Center for Education Statistics (NCES) staff, a policy advisory panel, and scholars in the field of interest.

Objectives developed and adopted by the Governing Board as a result of this process lead to NAEP assessments that are valid and reliable and that are based on widely accepted professional standards. The framework publications for each of the NAEP assessments provide more details about the development process for individual subjects. Frameworks are available at the Governing Board's website (http://www.nagb.org).

Frameworks differ from both the national and state content standards. While the standards document usually encompasses all that should be taught, the frameworks define only that which will be tested.

Nevertheless, the frameworks attempt to capture a broad range of content and skills that students need to learn in specific subject areas. The collaborative process used to develop the frameworks ensures that they reflect current educational requirements in a particular subject-area field.

Because the assessments must remain flexible to mirror changes in educational objectives and curricula, the frameworks must be responsive to current teaching practices and research findings. To ensure the currency of NAEP assessments, the frameworks are periodically revised so that test **specifications** still meet expectations of what students should know and be able to do in specific subject areas.

Specification of Assessment Questions

In addition to the framework, the Governing Board provides more detailed assessment specifications that guide item development. These specifications indicate how to implement and operationalize the intent of the framework.

Under the direction of NCES, current NAEP contractors develop the questions and tasks based on the subject-specific frameworks. National, state, and urban district main NAEP assessments use the same assessment instruments.

For each subject area assessment, a national committee of teachers, subject-matter specialists, and measurement experts provide guidance and review the questions to ensure that they meet the framework specifications. For each state assessment, state curriculum and testing directors review the questions to be included in the NAEP state component.

Related Questions:

Question 1: *What is NAEP?*

Question 4: *Why are NAEP questions kept confidential?*

Question 5: *How many schools and students participate in NAEP, and who are they? When are the data collected during the school year?*

Question 9: *What process is used to develop the assessments?*

question 2

Q: *Can the public examine the NAEP questions and find out how well individual students performed on the NAEP assessment?*

A: Most of the questions used in National Assessment of Educational Progress (NAEP) assessments remain secure or confidential to protect the integrity of the assessment. In order for NAEP to accurately measure student achievement over time, the assessments must be administered to students who have never seen the questions before. Nevertheless, NAEP typically stops using and releases about one-fourth of the questions used in each assessment. The released items are replaced with new items.

Released assessment questions may be viewed using a web-based tool on the NAEP website (http://nces.ed.gov/nationsreportcard/itmrls). This website also provides sample student booklets for the public to view or download.

Under certain prearranged conditions, small groups of people can also review the actual booklets being used in the assessment. This review must be arranged with the NAEP State Coordinator, with the National Center for Education Statistics (NCES), or with the National Assessment Governing Board. The review occurs under the supervision of NAEP program staff. The principal of a participating school can provide information about how to contact the State Coordinator for this purpose.

NAEP does not provide scores for individual children or schools since no individual student takes the entire NAEP assessment in a particular subject area. Each student instead answers a small subset of the entire assessment, which cannot accurately demonstrate a student's knowledge of a subject.

Further Details

Public Access to NAEP Questions

There are a number of ways in which the public can view the types of questions that NAEP will be asking students. The NAEP website (http://nces.ed.gov/nationsreportcard) provides parents, students, and others with sample test information (called Sample Questions Booklets) for downloading and printing.

In addition, NAEP has designed the NAEP Questions Tool, which provides web-based access to released questions in mathematics, reading, science, writing,

U.S. history, economics, civics, and geography. This tool allows the public to search for questions by grade (4, 8, and 12), by age (9, 13, and 17), by content area cognitive dimensions, by question type (i.e., **multiple-choice**, short constructed response, or extended constructed response), and by level of difficulty. The tool gives individuals an opportunity to see the NAEP questions, scoring guides/answer keys, sample student responses, overall student performance, and NAEP **student group** performance (e.g., performance by gender or race/ethnicity). A print component of the tool allows for easy printing of any combination of the released NAEP questions and ancillary material. The

NAEP Questions Tool is located online at http://nces.ed.gov/nationsreportcard/itmrls.

Within the limits of staff and resources, school administrators and interested members of the public can also make plans to view the actual NAEP **booklets** being used for the current assessment. Arrangements for this review must be made prior to the local administration dates so that sufficient materials can be provided and interested persons can be notified. Upon request, NAEP staff will also review the booklets with small groups of individuals in a secure setting, with the understanding that no assessment questions will be duplicated, copied, or removed.

While the public may examine the assessment questions, it is important to remember that NAEP does not provide scores for individual students or schools. To reduce the test-taking burden, no individual student takes the entire NAEP assessment in a particular **subject area**; rather, each student answers a small subset of the entire assessment. This subset is too small to provide an accurate picture of a particular student's knowledge of a subject. Therefore, it is not possible for NAEP to report results of an individual's performance. Instead, NAEP provides results for populations of students (e.g., fourth-graders) and subgroups of those populations (e.g., female students or Hispanic students).

Individuals who would like to view secure NAEP questions and instruments should

- make their request in writing;

- provide their name, affiliation, address, and telephone number; and

- direct their request to NCES, the Governing Board, or a NAEP State Coordinator.

NAEP State Coordinators have primary responsibility for coordinating with NCES to make arrangements for individuals to have access to secure NAEP questions and instruments. Contact information for the appropriate State Coordinator is available in NAEP state profiles or, for schools participating in NAEP, via the My NAEP website (http://www.mynaep.com). Contact information may also be obtained by calling NCES at 202–502–7420.

Related Questions:

Question 4: *Why are NAEP questions kept confidential?*

Question 16: *Are the NAEP data confidential?*

Q: *Why are NAEP questions kept confidential?*

A: As with other school tests or assessments, most of the questions used in the National Assessment of Educational Progress (NAEP) remain secure or confidential to protect the integrity of the assessment. For NAEP to accurately measure student achievement over time, the assessments must be administered to students who have never seen the questions before. Despite these concerns, NAEP typically releases one-fourth or more of the questions used in each assessment, making them available for public use.

Further Details

The Importance of Security

Measuring student achievement and comparing students' scores from previous years requires reusing some questions for continuity and statistical purposes. These questions must remain secure to assess trends in academic performance accurately and to report student performance on existing NAEP score scales.

Furthermore, for NAEP to regularly assess what the nation's students know and can do, it must keep the assessment from being compromised. If students have prior knowledge of test questions, then schools and parents will not know whether their performances are based on classroom learning or coaching on specific assessment questions.

Nevertheless, the public has a right to know about the content of NAEP assessments. NAEP stops using and releases to the public approximately 25 percent or more of the questions in each assessment after each assessment cycle, while maintaining the security of other NAEP questions for use in future assessments. These released questions are available to the public via the NAEP Questions Tool on the NAEP website, as described on page 10. Since NAEP has been assessing core **subject areas** and reporting trend data for subjects such as reading and mathematics since the early 1990s, the website contains a large collection of questions that represents the full range of grade levels assessed, question types, and the content classifications as specified by the subject-area **frameworks**.

Related Questions:

Question 3: *Can the public examine the NAEP questions and find out how well individual students performed on the NAEP assessment?*

Question 16: *Are the NAEP data confidential?*

Q: *How many schools and students participate in NAEP, and who are they? When are the data collected during the school year?*

A: The number of students selected to be in a National Assessment of Educational Progress (NAEP) sample depends on whether it is a national-only sample or a combined state and national sample. Generally, national assessments involve participation by fewer students and schools than state-level assessments. In the national-only sample, there are approximately 10,000 students per subject area and grade level. In a combined national and state sample, there are approximately 3,000 students sampled per participating jurisdiction from approximately 100 schools, per subject area and grade level. Typically, 30 students per subject per grade are randomly selected in each school.

Data for the national and state NAEP are collected at the same time during winter. While the best time for data collection may be the end of the school year when students have had more opportunity to learn, many states conduct their state assessments in the spring. By testing in the winter, NAEP interferes less with state programs. Data for the national long-term trend assessments are collected in the fall for 13-year-olds, in the winter for 9-year-olds, and in the spring for 17-year-olds. Other NAEP special studies can occur at different points throughout the school year.

Further Details

Sample Selection

NAEP does not, and is not designed to, report on the performance of individual students. Rather, it **samples** and reports on the performance of groups of individuals whose aggregate scores represent the performance of large **student groups**.

A sample is a subset of a population that is selected by an appropriate probability mechanism for the purpose of investigating the properties of a particular population. The total number of children in any particular grade in the United States is between three and four million. The number of students selected to be in a NAEP sample depends on whether it is a national-only sample, or a combined state and national sample (as would be the case for **subject areas** that are assessed at the state level). For subjects that are assessed at the national and state levels, approximately 4 percent of the students are sampled, including representative samples from each state. All the students from the combined sample comprise the national sample. For subjects that are assessed at the national level only, approximately 0.4 percent of the students are sampled to represent the entire population of U.S. students in the appropriate age or grade group. Different samples of the population of students at ages 9, 13, and 17 are selected for the NAEP long-term trend assessment.

Although only a very small percentage of the student population in each grade is assessed, NAEP estimates are accurate because they depend on the absolute number of students participating, not on the relative proportion of students. Thus, all or nearly all of the schools and students selected must partici-

question

pate in the assessment to ensure that the NAEP sample is truly representative of the nation's student population.

Ensuring Representative Samples

As the Nation's Report Card, NAEP must report accurate results for populations of students and subgroups of these populations (e.g., minority students or students attending nonpublic schools). The relatively small samples of students selected for the NAEP assessments must be truly representative of the entire student population.

Every school has a predictable chance of being selected for the sample. Factors such as grade, subject, or public and nonpublic status influence the probability of both school and student selection. Within a selected school, all students in a participating grade have equal likelihood of being chosen for the sample. However, the validity of statistically selected samples can be compromised by factors such as absenteeism or insufficient school participation. Conversely, the elective participation by unsolicited schools that do not fit the sampling design would undermine the validity of findings; therefore, while NAEP encourages the participation of all parties selected, it cannot accept volunteers.

A multistage design that relies on **stratification** (i.e., classification into groups having similar characteristics) is used to choose samples of student populations. To ensure an accurate representation, the samples are randomly selected from groups of schools that have been stratified by variables such as extent of urbanization, percentage of minority enrollment, median household income, or state achievement test results.

A nationally representative sample includes students from both participating and nonparticipating jurisdictions. Participating jurisdictions receive separate reports; students from nonparticipating jurisdictions form part of the national sample, but such jurisdictions do not receive separate reports.

For the national-only and long-term trend NAEP assessments, the sampling design begins with the selection of geographic areas defined as counties or groups of counties—termed **primary sampling units (PSUs)**. Then schools (public and nonpublic) are selected within the PSUs. Finally, students are selected within those schools. Stratification ensures that the PSU sample is representative of the nation.

To ensure that the results reported for major student groups of populations are accurate, **oversampling** (i.e., sampling particular types of schools at a higher rate than they appear in the population) is necessary. For example, for national-only assessments, main NAEP oversamples nonpublic schools and schools with large minority populations, thereby providing large samples to ensure accurate estimates of student group performance.

If these samples are to be representative of the population as a whole, however, the data from the students in the oversampled schools must be properly weighted during analysis. Weighting compensates for the disproportionate representation of certain student groups that occurs because of oversampling. Similarly, it also offsets low sampling rates that can occur for very small schools. Thus, when prop-

erly weighted, NAEP data provide results that reflect the representative performance of the entire nation and of the student groups of interest.

Assessment Schedules

NAEP does not assess all subjects at all grades every year. The independent National Assessment Governing Board, following the general requirements of federal legislation, determines which assessments will be assessed in particular years. Further information about assessment schedules for specific subjects is shown on page 55.

Within a particular assessment year, the most active period for NAEP assessments is the winter months. The time of year for conducting the assessment remains relatively constant across assessment years to permit an accurate measurement of change over time and to help ensure that the results are comparable.

National and state assessments, with the exceptions of arts and foreign language assessments, are administered during a 6-week period from the last week of January through the first week of March. Data collection activities for the long-term trend assessments occur in the fall for 13-year-olds, in the winter for 9-year-olds, and in the spring for 17-year-olds.

Related Questions:

Question 1: *What is NAEP?*

Question 2: *What subjects does NAEP assess? How are the assessment questions determined?*

Question 6: *How does NAEP use a large number of test questions, yet limit testing time per student to less than an hour?*

question 5

Q: *How does NAEP use a large number of test questions, yet typically limit testing time per student to less than an hour?*

A: Typically, several hundred questions are needed to reliably test the many specifications of the complex frameworks that guide the National Assessment of Educational Progress (NAEP) assessments. Administering the entire collection of subject-area questions to each student would be far too time consuming to be practical.

Therefore, NAEP divides the test questions into different portions, or blocks, and administers the various blocks of the entire pool of subject-area questions to different but equivalent student samples. This design minimizes the assessment time required per student while allowing complete coverage of the subject being assessed. NAEP assessments including background questions are designed so that they require approximately 90 minutes. Principals and teachers are asked to complete questionnaires--either online or on a paper copy. Teachers may also be asked to fill out questionnaires for their English language learners and students with disabilities.

NAEP asks each student to answer questions in only one subject and uses 20 to 60 varying combinations of different blocks from the item pool. This enables NAEP to check for any unusual interactions that may occur between different samples of students and different sets of assessment questions. NAEP distributes the test booklets in a way that ensures the different test forms are distributed in approximately equal numbers to each group of students in the sample.

Further Details

Design of NAEP Test Forms

In the NAEP design of test forms, the subject-area **blocks** are balanced. Each block of questions appears an equal number of times in every possible position in the various test **booklet** forms. Each subject-area block is also paired with every other subject-area block in at least one of the test forms. (The NAEP test form design varies according to **subject area**.) The number of subject-area blocks vary from 6 to 20, while the range of booklets goes from 18 to 73.

The following table presents a simplified example of **Balanced Incomplete Block (BIB) spiraling**. In this example, the full **sample** of students is divided into 15 equivalent groups, and each group of students is assigned one of the 15 test booklets. In this design, each subject-area block appears an equal number of times throughout all booklets (five times in this case). Each block is paired once with every other block. Each block appears two times in one booklet position and three times in the other position. (This example shows only the subject-area blocks, even though the test booklets also contain background questionnaire blocks.)

NAEP's test form design necessitates printing a greater variety of test booklets. Furthermore, each assessment booklet form must appear in the sample approximately the same number of times and must be administered to equivalent **student**

A Model of NAEP Test Forms

Booklet version	Position 1 subject-area block	Position 2 subject-area block
1	A	B
2	B	C
3	C	D
4	D	E
5	E	F
6	F	A
7	A	C
8	B	D
9	C	E
10	D	F
11	E	A
12	F	B
13	A	D
14	B	E
15	C	F

groups within the full sample. To ensure proper distribution at assessment time, the booklets are packed in order (in the above example, one each of booklets 1 through 15, then 1 through 15 again, and so on). The test coordinator randomly assigns these booklets to the students in each test administration **session**. Spiraled distribution of the booklets promotes comparable sample sizes for each version of the booklet, ensures that these samples are randomly equivalent, and reduces the likelihood that students will be seated within viewing distance of another student with an identical booklet.

Related Question:

Question 9: *What process is used to develop the assessments?*

Q: *What are NAEP's procedures for collecting data?*

A: Contractor staff administer the NAEP assessments after undergoing extensive training. Detailed procedural manuals, training, supervision, and quality control monitoring ensure uniformity of procedures across jurisdictions. The careful control of the complex data collection process contributes to the quality of the assessments and their results as well as ensuring that student and school information remains confidential.

Further Details

Organization and Supervision of Data Collection

The National Assessment of Educational Progress (NAEP) relies heavily on the participation of school administrators and staff. Obtaining the agreement of the selected schools requires substantial time and energy. A series of mailings, including letters to the chief state school officers and district superintendents, notifies the **sampled** schools of their selection. Additional informational materials are sent and procedures are explained at introductory meetings.

The data collection contractor is responsible for the following field administration duties:

- selecting the sample of schools and students;

- developing the administration procedures, manuals, and materials;

- hiring and training staff to conduct the assessments; and

- conducting an extensive quality-assurance program.

To meet the varying staffing demands of data collection for national, state, and long-term trend assessments, the contractor hires and trains between 1,000 and 3,000 field staff members every year. Field staff complete all NAEP-associated paperwork, reducing the burden on participating schools.

State supervisors work with state-appointed coordinators to carry out the necessary organizational tasks. The individual schools are responsible for preparing lists of students enrolled in the sampled grade, and distributing the teacher, school, and SD and/or ELL questionnaires. (SD and/or ELL refers to students with disabilities and/or English language learners.) NAEP contractor staff administer the assessment.

After each **session**, field staff interview school personnel to receive their comments and recommendations. As a final quality control step, the contractor solicits feedback from state personnel and from its own field staff to help improve procedures, documentation, and training for future assessments.

Management of Assessment Materials

Under the direction of the National Center for Education Statistics (NCES), the materials contractor produces the materials needed for the NAEP assessments. The contractor prints identifying bar codes and numbers for the **booklets** and questionnaires, preassigns the booklets to testing sessions, and prints the booklet numbers on the administration schedule.

These activities improve the accuracy of data collection and assist with the booklet distribution process. In order to ensure confidentiality, test booklet numbers are preassigned to the students in a particular assessment session; these numbers are printed on the administration schedule in advance of the testing date. Each student's name is recorded on a sticker temporarily affixed to the test booklet. Name stickers are removed and destroyed by the test administrator immediately after each session. Furthermore, the administration forms are perforated so that all student and teacher names can be easily removed after the administration session. At this point, all links between students' names and corresponding student, teacher, or school background information have been broken. The portion of the forms containing the student names is held by school administrators and destroyed on a pre-determined later date.

The materials contractor handles all receipt control, data preparation, and processing, scanning, and scoring activities for the NAEP assessments. Using an image-processing and scoring system specially designed for NAEP, the contractor scans the **multiple-choice** selections, the handwritten student responses, and other data provided by students, teachers, and administrators. When this **image-based scoring** system was introduced during the 1994 assessment, it virtually eliminated paper handling during the scoring process. The system also permits online monitoring and recalibration for scoring reliability.

Related Questions:

Question 3: *Can the public examine the NAEP questions and find out how well individual students performed on the NAEP assessment?*

Question 6: *How does NAEP use a large number of test questions, yet limit testing time per student to less than an hour?*

Question 10: *How does NAEP reliably score and process millions of student-composed responses?*

Question 15: *Who evaluates and validates NAEP?*

7

question

Q: *How does NAEP accommodate students with disabilities and English language learners?*

A: Throughout its history, the National Assessment of Educational Progress (NAEP) has encouraged the inclusion of all students who could meaningfully participate in the assessment, including special-needs students such as students with disabilities and/or English language learners. Over the years, schools have classified an increasing proportion of students as disabled (SD) and/or English language learners (ELL). Although NAEP establishes guidelines for inclusion, states differ in the types and levels of accommodation provided for SD and/or ELL students. Since the 1997 amendments to the Individuals with Disabilities Education Act (IDEA), however, some states are changing their criteria for including students with disabilities.

Previously, because of concerns about standardized administration, accommodations such as bilingual booklets and extended testing time were not permitted. As a result, some students who could have participated had accommodations been made available were excluded. In 1996 the National Center for Education Statistics (NCES) formally tested new inclusion policies for NAEP. Under these new, more inclusive guidelines, school administrators were encouraged, even when in doubt, to include SD and/or ELL students. In addition, the NAEP program began to explore the use of accommodations for these special-needs students. Based on analyses of the impact of offering accommodations, in 1996 NAEP began reporting results for some subject areas that included the performance of special-needs students who had received accommodations. Beginning in 2002, NAEP began reporting results for all subject areas that include the performance of accommodated students.

Further Details

Assessing Students With Special Needs

NAEP intends to assess all students selected to participate. However, some students may have difficulty with the assessment as it is normally administered because of a disability and/or because he or she is an English language learner. When a school identifies a student as having a disability and/or as being an English language learner, the teacher or staff member who is most familiar with the student is asked to complete a questionnaire about the services received by the student.

The anonymous **SD/ELL questionnaire** provides useful information about exclusion rates by disability conditions in different states. Students who cannot meaningfully take part, even with an accommodation allowed by NAEP, are **excluded** from the assessment. The decision to exclude SD and/or ELL students is made by local schools. They are encouraged to follow guidelines provided by the NAEP program.

Beginning with the 1996 national main assessment, NAEP implemented a two-part modification of procedures to increase inclusion in NAEP assessments. First, revised criteria were developed to guide how

decisions about inclusion should be made. Second, NAEP began providing certain accommodations that were either specified in a student's **Individualized Education Program (IEP)** or had been frequently used to test the student.

The accommodations vary depending on the subjects being assessed. Certain accommodations are not offered in particular **subject areas** if the use of the accommodations would alter the nature of the skills being assessed. For example, oral reading of the assessment passages and questions is not permitted for students participating in the NAEP reading assessment, and calculators are not allowed on parts of the NAEP mathematics assessment.

The following are examples of the accommodations that have been provided most frequently to students participating in NAEP:

- one-on-one testing;

- bilingual books in mathematics;

- large-print books;

- small-group testing;

- extended time;

- oral reading of directions;

- translating directions into American Sign Language;

- use of magnifying equipment;

- use of an aid for transcribing responses; and

- English-Spanish translation dictionary (except in the reading assessment).

In assessments conducted between 1996 and 2000, a **split-sample design** was used. The split-sample design made it possible to study the effects on NAEP results of including special-needs students who required and were provided with accommodations, while at the same time obtaining results that were comparable to those from previous assessments in which accommodations were not provided. Based on research conducted and published since that time, it was determined that NAEP could begin a transition to reporting results that included the performance of special-needs students who were assessed with accommodations. Beginning with the 2002 assessment, all students who require accommodations permitted by NAEP are allowed to use them.

Related Question:

Question 5: *How many schools and students participate in NAEP, and who are they? When are the data collected during the school year?*

Q: *What process is used to develop the assessments?*

A: To meet the nation's growing need for information about what students know and can do, the National Assessment of Educational Progress (NAEP) cognitive assessment instruments must meet the highest standards of measurement reliability and validity. Developing the assessment instruments—from writing questions to analyzing pilot-test results to constructing the final instruments—is a complex process that consumes most of the time during the interval between assessments. In addition to conducting national pilot tests, developers oversee numerous reviews of the assessment instrument by NAEP measurement experts, by the National Assessment Governing Board, and by external groups that include representatives from each of the states and jurisdictions that participate in the NAEP program.

Further Details

The Development Process

For each subject NAEP assesses, a subject-related **standing committee** is convened to provide input to the development process to help ensure that the assessment is aligned with the **framework** developed by the Governing Board. The subject-related standing committee reviews the assessment questions and independently confirms the validity of each question. The committee meets several times during the development cycle to consider how questions should be formatted, to verify grade appropriateness, to ensure usefulness for measuring subject-related knowledge or skills, to refine the scoring guides that will be used for scoring students' responses to **constructed-response questions**, and to review pilot-test results.

In addition to reviews by the subject-related standing committee, each newly developed assessment question is reviewed by National Center for Education Statistics (NCES) staff and approved by the Governing Board's Assessment Development Committee.

Furthermore, the assessments that are used in the state NAEP are reviewed by a group of state representatives. General assessment development issues are also discussed with a group composed of state representatives who meet regularly to consider issues related to the NAEP state assessment program.

The following summarizes the general steps used to develop the cognitive instruments for all NAEP assessments:

- Test development specialists and various subject-matter experts write the questions and exercises according to question **specifications** based on the frameworks for each subject.

- Test development staff experienced in the **subject area** review the questions and exercises for content concerns and revise them accordingly.

- Questions and exercises are put in a database, as is all the information that describes what the item is designed to test.

- Test developers assemble **blocks** of questions and exercises for national **pilot tests** according to specifications outlined in the subject frameworks. (NAEP uses the term "block" to refer to a collection of questions administered to students as a timed unit.)

- Specialists review the blocks for fairness, in order to eliminate potential item **bias**. At this time, copyright permission is obtained as necessary for any questions or exercises containing reprints of authentic source materials (such as reading passages or primary historical documents).

- Assessment questions are administered to individual students in one-on-one or small-group question tryout **sessions** to determine both how well students understand the questions and what further refinements should be made to the wording or formatting of questions.

- Subject-related standing committees are convened again to review the questions and blocks and to independently confirm that the questions fit the framework specifications and are correctly classified.

- For the state assessment program, assessment and curriculum specialists from participating states and other jurisdictions review all questions, exercises, and questionnaires that will be included in the assessment.

- Test developers update the pilot-test version of the questions and exercises based on reviews from the standing committee as well as content and assessment experts.

- The pilot-test questions are reviewed by NCES for compliance with government policies on data collection.

- The questions are then further reviewed by the Governing Board, which approves their use in the pilot test.

- The pilot tests are administered, scored, and analyzed.

- Suitable questions for the final assessment instrument are selected based on pilot-test results and framework specifications.

- Subject-matter specialists review the items selected for the final assessment.

- Assessment questions undergo additional fairness and editorial reviews.

- Subject-related standing committees are convened again to review the questions and to independently confirm **multiple-choice** answer keys, scoring guides, and classification codes.

- The final assessment questions are reviewed once again by NCES. The Governing Board further reviews these questions and revisions are made as needed to obtain government clearance from the Office of Management and Budget (OMB).

- The assessments are administered, scored, and analyzed.

question 9

question 9

The blocks undergo a mandatory fairness review to ensure that the assessment reflects thoughtful, balanced input from all groups of people. External reviewers, including state education agency personnel, review the questions for appropriateness for students from a variety of backgrounds and across **regions**. As part of its responsibility for final approval of all NAEP assessment questions, the Governing Board ensures that all questions selected for NAEP are free from racial, cultural, gender, or regional bias and are nonideological, secular, and neutral.

After assessments are conducted, the results for each assessment question are checked empirically. This empirical check for fairness employs **differential item functioning (DIF)** analyses. DIF analyses identify questions that are differentially difficult for particular **student groups** (identified by categories such as racial/ethnic classification or by gender) for reasons that seem unrelated to the overall ability of the students. For further discussion of procedures for detecting DIF, see the The NAEP 1998 Technical Report (Allen, Donoghue, and Schoeps, 2001).

Related Questions:

Question 2: *What subjects does NAEP assess? How are the assessment questions determined?*

Question 11: *How does NAEP analyze the assessment results?*

A: National Assessment of Educational Progress (NAEP) assessments contain both multiple-choice and constructed-response questions. While multiple-choice questions allow students to select an answer from a list of options, constructed-response questions require students to provide their own answers. Whereas responses to multiple-choice questions are scored by a computer scoring program, responses to constructed-response questions are scored by qualified and trained scorers.

Scoring a large number of constructed responses with a high level of reliability and within a limited time frame is essential to NAEP's success. (In a typical year, over three million constructed responses are scored.) To ensure reliable, quick scoring, NAEP takes the following steps:

- develops focused, explicit scoring guides that match the criteria delineated in the assessment frameworks;

- recruits qualified and experienced scorers, trains them, and verifies their ability to score particular questions through qualifying tests;

- employs an image-processing and scoring system that routes images of student responses directly to the scorers so they can focus on scoring rather than paper routing;

- monitors scorer consistency through ongoing reliability checks;

- assesses the quality of scorer decision making through frequent monitoring by NAEP assessment experts; and

- documents all training, scoring, and quality control procedures in the NAEP technical reports.

Further Details

Developing Scoring Guides

Scoring guides for the assessments are developed using a multistage process. First, scoring criteria are articulated. While the **constructed-response** questions are being developed, initial versions of the scoring guides are drafted. Subject-area and measurement specialists, the subject-related **standing committees**, the National Center for Education Statistics (NCES), and the National Assessment Governing Board review the scoring guides to ensure that they include criteria consistent with the wording of the questions; are concise, explicit, and clear; and reflect the assessment **framework** criteria.

Next, the guides are used to score student responses from the **pilot test**. The subject-related standing committees and contractor staff use pilot-test results to further refine the guides. Finally, training materials are prepared. Assessment specialists select examples of student responses from the actual assessment for each performance level specified in the guides. Selecting the examples and **anchor sets** provides a final opportunity to refine the wording in the scoring guides, develop additional

question

training materials, and make certain that the guides accurately represent the assessment goals set forth in the framework.

The student response examples clearly express a committee's interpretations of each performance level described in the scoring guides and help to illustrate the full range of achievement under consideration. Further, the examples promote consistent interpretation of scoring guides during the actual scoring process, helping to ensure the accurate and reliable scoring of diverse responses.

Recruiting and Training Scorers

Recruiting highly qualified trainers and scorers to evaluate students' responses is crucial to the success of the assessment. A four-stage model is used for selecting and training scorers.

The first stage involves selecting scorers who meet qualifications specific to the **subject areas** being scored. Prospective scorers participate in a simulated scoring exercise and a series of interviews before being hired. (Some applicants—particularly those who will be scoring the mathematics, reading, science, and writing assessments—take an additional exam to measure their understanding of specific skills.)

Next, scorers are oriented to the project and trained to use the **image-based scoring** system. This orientation includes a presentation of the goals of NAEP and the frameworks for the assessments.

Preparing Training Materials

Training materials, including sample student responses, are then prepared for the scorers. Trainers and scoring supervisors read hundreds of student responses to select sample responses that represent

each level in the scoring criteria. The samples are selected to ensure representation of students according to the following categories: the different types of schools participating in the assessment; race/ethnicity; gender; geographical location; and by **region** of the country.

In the third stage, subject-area specialists train scorers using the following procedures:

- presenting and discussing the exercise or question to be scored and the scoring rationale;

- presenting the scoring guide and the sample responses;

- discussing the rationale behind the scoring guide, with a focus on the criteria that distinguish the various levels of the guide;

- practicing the scoring of a common set of sample student responses known as anchor papers;

- discussing in groups each response contained in the practice scoring set; and

- continuing the practice steps until scorers reach a common understanding of how to apply the scoring guide to student responses.

In the final stage, scorers assigned to extended constructed responses work through a qualification round of sample student responses to ensure accuracy and consistency in applying the scoring guide. At every stage, NAEP staff closely monitor scorer selection, training, and quality.

Using the Image-Based System

The image-based scoring system was designed to accommodate NAEP's specific needs while eliminating many of the complexities involved in paper-based training and scoring. First used in the 1994 assessment, the image-based scoring system allows scorers to assess and score student responses on a computer. To do this, student response **booklets** are scanned, constructed responses are digitized, and the images are stored for presentation on computer monitors. The range of possible scores for an item also appears on the display, so scorers can quickly click on the appropriate button to register their scores.

The image-based system facilitates the training and scoring process by electronically distributing responses to the appropriate scorers and by allowing NAEP supervisors to monitor scorer activities, identifying problems as they occur and implementing solutions expeditiously.

The image-based scoring system allows for all student responses to a single question to be scored continuously, rather than scoring individual student booklets containing responses to multiple questions. This grouping of all student responses to each question improves the validity and reliability of scorer judgments.

Ensuring Rater Reliability

Rater reliability refers to the consistency with which individual scorers assign the same score to a constructed response. This consistency is critical to the success of NAEP; therefore, project staff employ three methods for monitoring reliability.

In the first method, called "backreading," scoring supervisors selectively review each scorer's work to confirm that the scorer applies the scoring criteria accurately and consistently over time and across a large number of responses. At least 5 percent of each scorer's work is monitored in this process.

In the second method, each group of scorers performs calibration as needed throughout scoring, enabling supervisors to monitor and prevent scoring drift. After scorers have taken an extended break (e.g., at the start of the workday, after lunch), they review the scoring guide and training set and may score a **calibration set** of papers to reinforce the scoring criteria before returning to score actual student responses.

Last, interrater reliability statistics confirm the degree of consistency in overall scoring, which is measured by scoring a defined percentage of the responses (5% for state assessments, 25% for national assessments) a second time (by a second, different scorer) and comparing the first and second scores.

Maintaining Scoring Consistency

Consistent performance among scorers is paramount for the assessment to produce meaningful results. NAEP's scoring contractors have designed the image-based scoring system to allow for easy monitoring of the scoring process, early identification of problems, and flexibility in training and retraining scorers.

Measuring trends in student achievement, whether short or long term, involves special scoring concerns. To compare student performance across years, scorers must train using the same materials and procedures as in previous assessment years. Furthermore, interrater reliability rates and item mean score drift must be

question 10

monitored within the current assessment year as well as across years.

To maintain scoring consistency across years, a random **sample** of approximately 2000 responses to each question from the prior assessment is randomly interspersed among current responses for rescoring; approximately 500 additional responses are used for trend training. The results are used to determine the degree of scoring agreement between the current and previous assessments.

Documenting the Process

The NAEP Technical Documentation is written for researchers familiar with educational measurement and testing and can be accessed online (http://nces.ed.gov/nationsreportcard/tdw). Users will find information concerning item development; the content chosen to be assessed; instruments used in the NAEP assessments; accommodations made for students with disabilities; and the NAEP database, which contains assessment information collected from students and teachers. The database does not contain identifying information and is intended solely for statistical purposes.

Related Questions:

Question 11: *How does NAEP analyze the assessment results?*

Question 13: *How does NAEP make reports and information available to the public?*

Q: *How does NAEP analyze the assessment results?*

A: Before the data are analyzed, responses from the subgroups of students assessed are assigned sampling weights to ensure that their representation in National Assessment of Educational Progress (NAEP) results matches their actual percentage of the school population in the grades assessed.

Then, data for national and state NAEP assessments in most subjects are analyzed by a process involving the following steps:

- Check item data and performance: The data and performance of each item are checked in a number of ways, including checks on scoring reliability and on differential performance by population groups that is unrelated to overall scores, to ensure fair and reliable measures of performance in the subject of the assessment.

- Set the scale for assessment data: Each subject assessed is divided into subskills, purposes, or content domains specified by the subject framework. For example, the 2009 reading assessment specifies three purposes for reading at grade 8, while the 2007 mathematics assessment specified five content domains, and the 2009 science assessment specifies three content domains. Separate scales are developed relating to the content domains in an assessment subject area. A statistical procedure, Item Response Theory (IRT) scaling, is used to estimate the measurement characteristics of each assessment question.

- Estimate group performance results: Because NAEP must minimize the burden of time on students and schools by keeping assessment administration brief, no individual student takes more than a small portion of the assessment for a given content domain. NAEP uses the results of scaling procedures to estimate the performance of groups of students (e.g., of all fourth-grade students in the nation, of female eighth-grade students in a state).

- Transform results to the reporting scale: Results for assessments conducted in different years are linked to reporting scales to allow comparison of year-to-year trend results for common populations on related assessments.

- Create a database: A database is created and used to make comparisons of all results, such as scale scores, percentiles, percentages at or above achievement levels, and comparisons between groups and between years for a group. All comparisons are subjected to testing for statistical significance, and estimates of standard errors are computed for all statistics.

To ensure reliability of NAEP results, extensive quality control and plausibility checks are carefully conducted as part of each analysis step. Quality control tasks are intended to verify that analysis steps have not introduced errors into the results. Plausibility checks are intended to encourage thinking about whether the results make sense and what story they tell.

question 11

Weighting

NAEP uses weights to ensure that student **samples** and subsamples are representative of their respective population groups. Each student assessed represents a portion of the population of interest. **Sampling weights** are needed to make valid inferences between the student samples and the respective populations from which they were drawn. Responses from the **student groups** are assigned sampling weights to adjust for **oversampling** or undersampling from a particular student group. For instance, in national-level-only assessments, census data on the percentage of Hispanic students in the entire student population are used to assign a weight that adjusts the proportion of Hispanic students in the NAEP sample to be nationally representative.

A statistician assigns a weight to each student that is the inverse (or reciprocal) of the student's **selection probability**. Since ignoring the fact that data cannot be assumed to be randomly missing could **bias** results, NAEP makes adjustments to weights to correct for detectable types of school-level and student-level **nonresponse**. When response rates are low, NAEP conducts analyses to assess the extent of possible biases that may have been introduced. All NAEP analyses described below are conducted using these nonresponse adjusted sampling weights.

Steps in NAEP Analysis

Check Item Data and Performance

A portion of the items on every NAEP assessment are **constructed-response** items, which require that the student create a response rather than select one from a provided set of choices. Such items require scoring by human raters. Lack of consistency between raters may reduce the reliability of the assessment results. To ensure the quality of within-year and across-year scoring reliability, statistical monitoring processes are implemented to assure that specific NAEP reliability standards are met. NAEP analysis staff and scoring staff are in regular communication about rating consistency issues in order to ensure that any scoring inconsistencies are resolved appropriately in a timely fashion.

All subject-area and **background questions** are subjected to an extensive quality control analysis. Project staff members review the item analysis results, searching for anomalies that may signal unusual results or errors in creating the database. Simultaneously, each subject-area question is examined for **differential item functioning (DIF)**. DIF analyses identify questions, if any, on which the scores of different subgroups of students, such as males and females, differ significantly after matching on ability level. Questions showing such differences are examined by experts for potential bias toward particular student subgroups.

Set the Scale for Assessment Data

After the item and DIF analyses have been completed, the **Item Response Theory (IRT)** scaling phase begins for each individual grade level and subject. NAEP uses IRT methods to produce a common scale for all assessment performance data (for the nation and all the states together), so scores and trends can be reported on a common metric. IRT scaling provides a method for summarizing

performance on all test questions that measure a common **content domain**. IRT scaling defines the common content by quantifying the relationships between the content scale and the assessment questions in terms of difficulty, discrimination, and other item parameters. Parameters of the IRT model are estimated for each question, with separate scales being established for each predefined content domain (a single scale within a **subject area**) specified in the assessment **framework**.

For example, the 2007 reading assessment for grade 8 had three scales describing reading purposes: reading for literary experience, reading to gain information, and reading to perform a task. Because the item parameters determine how each question is represented in the content domain scales, project staff employ **psychometric** methods to verify that the IRT scaling model provides an acceptable representation of the responses to the questions. In particular, they examine the fit of the model for each question. Item parameter estimation is performed on the entire sample of student responses to subject-area questions.

Estimate Group Performance Results

NAEP's basic goal is to report performance for groups of students on broad content and skill areas. NAEP's main interest is examining group statistics (such as average scale score, percentages of students at or above certain **achievement levels**, and percentiles) and comparing these statistics across groups (e.g., males vs. females) and over time (e.g., males in 2007

compared to males in 1996). In theory, given a sufficient number of questions in a content domain, performance distributions for any population could be determined for that content domain. However, NAEP must minimize its burden on students and schools by keeping assessment time brief. To do so, NAEP breaks up most assessments into approximately 10 **blocks**, each consisting of multiple questions, and administers 1 to 3 blocks of questions to any particular student, depending on the subject. As a result, any given student responds to only a small number of assessment questions for each content domain. Consequently, the performance of any particular student cannot be measured accurately. This student-level imprecision has two important consequences: first, NAEP cannot report the proficiency of any particular student in any given subject area; and second, traditional statistical methods that rely on **point estimates** of student proficiency become inaccurate and ineffective.

To resolve the apparent dilemma of imprecision in student-level measurement, NAEP uses methodology that produces estimates of the population distribution characteristics directly, without the intermediary stage of calculating point estimates for individuals. This is accomplished using the technique of marginal maximum likelihood estimation, meaning that NAEP scale score distributions are based on an estimated distribution of scale scores, rather than point estimates of a single scale score. This approach allows NAEP to produce accurate and statistically unbiased estimates of population characteristics that properly account for the imprecision in student-level measurement.

question 11

question 11

Transform Results to the Reporting Scale

After the group performance results have been estimated, the data are then linked to the reporting scale for the related assessments. The transformation to a trend reporting scale is done through a common population linking, which consists of the same students taking the same test analyzed two different ways. Over half the items administered in both years of adjacent assessments are identical. Item parameters for identical items are constrained to be equal in both the current and the previous assessment and re-estimated. Means and **standard deviations** are recalculated for the previous assessment with the new item parameters.

The overall mean and standard deviation of the previous assessment (as re-estimated in the current year with the joint IRT item parameters) are matched to the mean and standard deviation of the previous assessment using the original IRT item parameters through a linear transformation. The same linear transformation is then applied to the distribution of the current year's data. As a result, both years' data are comparably placed on the same reporting scale. Comparing the score distributions for population groups within the overall population determines the adequacy of the linking function.

Create a Database

Results, such as scale scores and percentiles, are compared using a database. A database is also used for creating comparisons between groups or between years for the same group. Statistical tests must be conducted to ensure that changes or differences between two numbers stem from dependable population differences and not sampling or measurement errors.

Statistical significance of NAEP results such as average scale scores, standard deviations, percentiles, percentages at or above achievement levels, and percentages of the population represented by groups are computed and reported. Since all NAEP statistics are subject to measures of uncertainty due to sampling error and measurement error, estimates of standard errors should also be computed to reflect the amount of uncertainty.

Related Question:

Question 13: *How does NAEP make reports and information available to the public?*

A: Researchers, policymakers and other interested parties can use the NAEP data provided by the National Center for Education Statistics (NCES) to perform their own analyses and studies on educational achievement. Additionally, NCES organizes seminars and discussions to address educational research questions using NAEP data at both the national and state levels.

Further Details

NAEP Data and Results

Because of its large scale, the regularity of its administration, and its thorough quality control process for data collection and analysis, NAEP provides numerous opportunities for secondary data analysis. NAEP data are used by researchers who have many interests, including educators who have policy questions and research scientists who study the development of abilities across the three grades assessed by NAEP.

NAEP has developed products that support the complete dissemination of both national and state NAEP results and data to many audiences. Key data about each state's or jurisdiction's schools and student population, as well as its NAEP testing history and results is located in the State Profiles section of the website. This section also offers links to other sources on the website, including the most recent state report cards for all available subjects, scale scores, **achievement levels**, and key instructional variables. These tools and more are found at http://nces.ed.gov/nationsreportcard.

NAEP Outreach

In addition to these products and tools, NCES periodically offers sem-inars to stimulate interest in using NAEP data to address educational research questions, enhance participants' understanding of the methodological and technological issues relevant to NAEP, and demonstrate the steps necessary for conducting accurate statistical analyses of NAEP data. In addition to offering formal and hands-on instruction, the seminars help participants learn about and work with currently available software packages specifically designed for NAEP analyses. These seminars are advertised in advance on the NCES website (http://nces.ed.gov/conferences).

NAEP also conducts discussions of educational issues and policies with state, district, and jurisdiction representatives. Participants in these discussions include testing directors, NAEP coordinators from individual states and other jurisdictions, and representatives from nonpublic school organizations and associations. NAEP also offers information about upcoming assessments and enables those involved in state NAEP to offer their input.

Related Question:

Question 13: *How does NAEP make reports and information available to the public?*

Q: *How does NAEP make reports and information available to the public?*

A: The National Assessment of Educational Progress (NAEP) has developed a number of different publications and web-based tools that provide direct access to national and state data and information. NAEP produces printed reports that offer a comprehensive view of student achievement in particular subject areas. In addition, NAEP has increasingly leveraged the power of the Internet to disseminate assessment results and reports.

NAEP's websites (http://nces.ed.gov/nationsreportcard, http://nationsreportcard.gov) provide more than just access to printed reports; they house a number of important web-based applications that deliver comprehensive NAEP data and information to the public. There are web pages that highlight results for every major NAEP release. In addition, NAEP has developed a web-based tool, the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata), that provides access to extensive NAEP results beyond what appears in print.

Further Details

NAEP Printed Reports

NAEP Report Cards comprehensively report all major results for each assessment. Overall performance results for the nation, states, and a few large urban school districts are offered, as well as the results of demographic **student groups** as defined by variables such as gender, race/ethnicity, type of school, school location, eligibility for free/reduced-price school lunch, and parents' highest level of education. In addition, other factors that can affect student performance, such as instructional activities and school policies, may be presented. These reports also provide relevant information on the development, scoring, and analysis of the assessment. Average scores, achievement-level results, percentages of students within defined student groups, and **standard errors** for all the data presented in the body of the report are available on the NAEP website within the Data Explorer.

Trial Urban District Assessment (TUDA) Reports provide a printed summary of results for selected large urban school districts.

Technical Reports document the **psychometric** details of the national and state assessments, including the **sample** design, instrument development, data collection process, and analysis procedures. Technical reports provide information about how the results of the assessment were derived; they do not present the actual results.

The NAEP Website

The NAEP websites (http://nationsreportcard.gov and http://nces.ed.gov/nationsreportcard) provide platforms for the dissemination of NAEP results, data, and general program information. For every major assessment release, web pages are created with graphics and text that highlight the results. Subject-specific pages explore how the NAEP assessments are developed, what they are intended to measure, and where users

can find the latest results and reports. In addition, the website houses important general information regarding the NAEP program and specific pages of information for those schools that are selected to participate in the NAEP assessment.

A unique aspect of the website is the presence of web-based tools that allow users to access NAEP questions, NAEP data, and state-specific NAEP information. Tutorials on the website guide users so they can effectively utilize the tools. Web products and applications are continually augmented and enhanced to maximize the effective dissemination of NAEP data and results.

The NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata) provides access to all NAEP data that have been collected since 1990. It provides users with direct access to NAEP national and state data, allowing users to generate and customize their own data tables and graphics. Users are able to create tabular and graphical representations of results and to download tables and graphics into commonly used software packages for personal use or presentations. Users can also perform significance tests to see if observed differences in data are **statistically significant**.

The NAEP Questions Tool (http://nces.ed.gov/nationsreportcard/itmrls/startsearch.asp) houses a database of released NAEP questions in the subjects that NAEP assesses. All three grade levels are represented, as are all question types (i.e., **multiple-choice** and **constructed response**). The tool allows users to search for questions by subject, grade, **framework** classification, question type, and

level of difficulty. Users then have access to NAEP questions, scoring guides/keys, sample student responses, overall student performance, and NAEP student group performance (e.g., gender, racial/ethnic, and achievement-level performance). A print component within the tool allows users to easily print any combination of NAEP questions and ancillary material.

The NAEP State Comparisons Tool (http://nces.ed.gov/nationsreportcard/nde/statecomp) provides data on student performance in mathematics, reading, science and writing assessments from each individual state and the District of Columbia. This tool allows users to create tables, sort data and compare states and jurisdictions based on the average scale scores for selected groups of public school students. Users can see how groups of students performed within a single assessment year or how performance has changed from a previous assessment year to the most recent.

The NAEP Item Maps (http://nces.ed.gov/nationsreportcard/itemmaps) presents examples of student performance and knowledge in NAEP **subject areas** at each achievement-level. Hyperlinked items allow users to view the item, scoring guide, answer key, student responses and performance data. These items tie into the NAEP Test Yourself and Questions Tools, allowing users to take an in-depth look at information presented to students taking the NAEP assessments. Items that are not hyperlinked are still in use and have not been released to the public.

The NAEP Test Yourself Tool (http://nationsreportcard.gov/testyourself.asp) gives users the opportunity to attempt to answer actual questions that have appeared in NAEP assessments. Questions are divided by subject area and grade level, allowing students, parents and other inter-

question 13

ested parties to try their hand at a variety of questions in both multiple-choice and constructed-response question types.

State Report Cards and District Snapshot Reports (http://nces.ed.gov/nationsreportcard/pubs/dst2005/2006458.asp) provide quick access to state- and district-level results and a history of state participation in the NAEP assessments. These pages also provide direct access to the NAEP Data Explorer to investigate the wealth of state and district data on the website.

Related Question:

Question 12: *How do NCES and members of the public work together to explore education issues using NAEP data and results?*

Q: *Can NAEP results be linked to other assessment data?*

A: In recent years, there has been considerable interest among education policymakers and researchers in linking National Assessment of Educational Progress (NAEP) results to other assessment data. Linking allows researchers to predict from students' performance on one assessment how they might perform on another assessment they did not take. The 1992 NAEP mathematics assessment results were successfully linked with the International Assessment of Educational Progress (IAEP) of 1991, and the 1996, 2000, and 2003 grade 8 NAEP assessments in mathematics and science have been less successfully linked to the Trends in International Mathematics and Science Study (TIMSS) of 1995, 1999 and 2003. Various methods for linking NAEP scores to state assessment results continue to be explored. Methods continue to be explored to enhance the value of NAEP data by linking to other national databases, such as the Common Core of Data and the School and Staffing Survey.

Further Details

Linking NAEP to International Assessments

The International Assessment of Educational Progress (IAEP).

Pashley and Phillips (1993) investigated linking mathematics performance on the 1991 IAEP to performance on the 1992 NAEP. In 1992, they collected sample data from U.S. students who were administered both instruments.

A **regression analysis** model was developed and then used for projecting IAEP scores from non-U.S. countries onto the NAEP scale.

The relation between the IAEP and NAEP assessments was relatively strong with a good model fit. However, the authors cautioned that linking of results should be considered only if two assessments are similarly constructed and scored.

Trends in International Mathematics and Science Study (TIMSS).

The results from the 1996 NAEP and the 1995 TIMSS assessments were linked by matching their score distributions (Johnson and Owen, 1998), since the two assessments were conducted in different years with no students taking both assessments. A comparison of linked eighth-grade results with actual eighth-grade results from states that participated in both assessments suggested that the link was working at an acceptably valid level.

The same linking approach produced inconsistent results at grade 4; therefore, no comparisons at this grade were reported. No studies have explained why the distribution matching method produced consistent results at only one grade.

TIMSS (2003).

Using **equipercentile equating**, 2003 NAEP data were linked to 2003 TIMSS data (Phillips, 2007) to estimate the percentage of eighth graders in each country that would perform at or above each of the NAEP **achievement levels**. The results showed that only Singapore and Taiwan had students whose average science score was equivalent to NAEP's science proficient level. In mathematics, Singapore, South Korea, Hong Kong, Taiwan and Japan students scored, on average, at NAEP's proficient level.

NAEP Scores and State Assessment Results

One way in which NAEP can be made most useful to state education agencies is by providing a benchmark for comparing the results of the local and state assessments conducted in their schools. If a state's assessment results show a similar pattern of improvement to the state's NAEP scores, conclusions about progress toward state education goals will be strengthened.

Linking NAEP Data with Other Databases.

Building on the earlier work of Linn (1993); Bloxom, Nicewander, and Tan (1995); and Williams et al. (1995), McLaughlin (1998a) explored the feasibility and validity of regression-based linking based on matching state assessment scores of students to NAEP performance records. Using the 1996 state NAEP grade 4 and 8 mathematics assessments in four states, he found (a) it is feasible to develop the linkage of student records without violating either NAEP or state assessment confidentiality assurances, and (b) in three of the four states, acceptably accurate

regression estimates of group-level NAEP scores and percentages at achievement levels could be obtained.

McLaughlin (1998b) found that in order for comparisons to be neutral (i.e., so that comparisons based on projected NAEP scores lead to the same conclusions as comparisons based on actual NAEP scores), state test values for average school scores and individual student scores, as well as demographic measures, must be included in the regression models. Like others (Linn and Kiplinger, 1993; Shepard, 1997), he also found that regression functions did not necessarily generalize across years.

Note that many factors influence the validity of inferences that can be drawn from linked scores. These factors include, but are not limited to, the content assessed, the format of the assessment items, the length of the assessment, and the amount of error present in the estimates. Unless the assessment to be linked to NAEP is very similar to NAEP on all of these factors, the linkage could be unstable and potentially misleading. If the test to be linked to NAEP differs from NAEP on any of these factors, some limited interpretations of the linked scores may be defensible, but others may not.

Braun (2007) and McLaughlin (2007) evaluated the 2005 NAEP as a common yardstick for comparing the proficiency standards each state sets on its own tests for fourth and eighth grade reading and mathematics, and for comparing these state standards with national performance benchmarks.

The findings show that states vary widely in the NAEP-equivalents of their proficiency standards. There is a 55 to 81-point difference in proficiency standards between the states, about twice the range seen in average student performance on NAEP between states. Most state proficiency standards fall within the NAEP Basic range—except in 4th-grade reading, where most fall below Basic. It should be noted that the NAEP definition of proficient "competency over challenging subject matter" is different than the states' definition. A state's proficiency standard is not necessarily tied to student performance on NAEP. For example, a state may have a less rigorous Adequate Yearly Progress standard, but consistently score highly on NAEP.

The 2007 NAEP reading results are currently in the process of being linked with the Educational Childhood Linking Study-Kindergarten cohort in an effort to conduct studies on informing the development of socioeconomic status measures for NAEP, and to estimate achievement growth curves for NAEP.

Related Question:

Question 12: *How do NCES and members of the public work together to explore education issues using NAEP data and results?*

question 14

15
question

Q: *Who evaluates and validates NAEP?*

A: Because National Assessment of Educational Progress (NAEP) findings have an impact on the public's understanding of student academic achievement, precautions must be taken to ensure the validity and reliability of these findings. Therefore, in its current legislation, as in previous legislative mandates, Congress has called for ongoing evaluation of the assessment as a whole. In response to these legislative mandates, the National Center for Education Statistics (NCES) has established various expert panels to study NAEP. These panels have produced a series of reports that address numerous important NAEP issues.

Further Details

Evaluation

A variety of organizations and individuals are continually involved in the evaluation of both the content and technical aspects of NAEP assessments. In the late 1980s and early 1990s, a Technical Review Panel (TRP) was convened by NCES to conduct a thorough evaluation of the NAEP program. The committee's white paper, Assessing the Validity of the National Assessment of Educational Progress: NAEP Technical Review Panel White Paper, recommended ongoing validation studies for the NAEP assessments (Linn, Koretz, and Baker, 1996). In addition, the National Academy of Education (NAE) was awarded a grant by NCES to evaluate both the state assessment program during its first few years of implementation (Glaser, Linn, and Bohrnstedt, 1997) and the National Assessment Governing Board's **achievement levels** (Shepard et al., 1993).

In recent years, evaluations have been conducted on an ongoing basis in two different ways. First, reviews and evaluations of the content of the NAEP assessments are conducted regularly by subject-related **standing committees** and by NCES and Governing Board staff. In addition, various Governing Board subcommittees are responsible for oversight of different aspects of the program. The Committee on Standards, Design, and Methodology monitors external contracts; the Committee on Reporting and Dissemination prepares and recommends procedures for reporting and disseminating NAEP results; and the Assessment Development Committee reviews and recommends test content for NAEP. Second, panels are formed periodically by NCES or external organizations such as the National Academy of Sciences (NAS) to conduct evaluations in accordance with congressional mandates.

In 1996, NAS was awarded a contract to further evaluate national and state NAEP. In response, NAS formed a committee of distinguished educators and other experts to conduct the evaluation activities described in the congressional mandate of 1994 Public Law 103–382, stating that "the Secretary shall provide for continuing review of the National Assessment, State Assessments, and student performance levels by one or more nationally recognized organizations." In the evaluation process,

the NAS committee directed work-shops, commissioned papers, solicited testimony and interviews, observed NAEP activities, and studied program documents, extant research, and prior evaluation reports. Based on this process, NAS released its NAEP evaluation report, Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress (Pellegrino, Jones, and Mitchell, 1999). The report presented observations and recommendations for a number of key functions, including (1) streamlining the design of NAEP; (2) enhancing the participation and meaningful assessment of English language learners (ELL) and students with disabilities (SD); (3) broadening the **framework** design and the assessment development process; and (4) setting reasonable and useful performance standards. The full text of the 1999 report is available online at the NAS website (http://books.nap.edu/ catalog.php?record_id=6296).

In 2005, the Buros Center for Testing, in collaboration with the University of Massachusetts/Center for Educational Assessment and the University of Georgia, was awarded the contract to conduct an external evaluation of NAEP.

The NAEP Validity Studies Panel

NCES established the NAEP Validity Studies (NVS) Panel to provide technical review of NAEP plans and products, to identify technical concerns and promising techniques worthy of further study and research, and to conduct small-scale validity studies.

Since its inception in October 1995, the NVS Panel has worked on numerous validity studies. The panel has released reports on topics such as assessment design, item format, assessment technologies, sampling, equating, and reporting assessment results. The released reports are available online at the NAEP Research E-Center website (http://nces.ed.gov/nationsreportcard/researchcenter/papers.asp).

Related Question:

Question 7: *What are NAEP's procedures for collecting data?*

question 15

16

question

Q: *Are NAEP assessment data confidential?*

A: The National Assessment of Educational Progress (NAEP) program undertakes measures to ensure the confidentiality of all schools and students who participate in the assessments. After publishing NAEP reports, the National Center for Education Statistics (NCES) makes the data available to researchers, but withholds student and school names and other identifying information. Although it might be possible for researchers who have received special access to data to deduce the identities of some NAEP schools, they are bound, under penalty of fines and prison terms, to keep these identities confidential.

Further Details

A Confidential Assessment

Detailed, codified test administration procedures assure the confidentiality of all students who take NAEP assessments. The names of students are used to assign specific test **booklets** to students selected for a particular assessment. Each booklet has a unique, temporary identification number so that it can be linked to teacher and school data. After a student completes the assessment, NAEP no longer needs students' names, and the links between students' names and their test booklets are destroyed by school administrators.

NAEP administrators use tear-off forms to break the link between the names and identification numbers before test booklets are sent for scoring and analysis. Before administrators send booklets to be scored, they remove the portion of the form containing the student's name. Local school officials keep these forms in a secure storage envelope for a few weeks after the assessment in case the link to the identification numbers needs to be checked. When the information is no longer needed, schools are notified and officials destroy

the storage envelope, confirming their actions by returning a Destruction Notice to NAEP. In addition, all government and contractor employees who work with NAEP data collection, analysis, and reporting swear to uphold a confidentiality law. If any employee violates the confidentiality law by disclosing the identities of NAEP **respondents**, that person is subject to criminal penalties.

Released Data

NAEP provides results about subject matter achievement, instructional experience, and school environment and reports these results for populations of students (e.g., fourth-graders) and subgroups of those populations (e.g., male students or Hispanic students). NAEP does not provide individual scores for the students or schools assessed.

In addition, the data that are released in published reports and on the NAEP website cannot be traced to any particular school or student. Under NCES confidentiality laws and supporting procedures, released data must be certified as clean, or purged of individually identifiable information, before being made available to the general public.

Education researchers may have an interest in additional analyses that require access to raw NAEP data. As a publicly funded project, NAEP fulfills the requirement to make such data available on a restricted-use basis by offering national and state data files to researchers. Qualified researchers interested in obtaining a Restricted-Use Data License, visit http://nces.ed.gov/statprog/instruct.asp for more information and an application.

Before releasing raw data, NCES requires that researchers agree to the terms of the Restricted-Use Data License, including a security plan, inspections for compliance, submission of releases for confidentiality review, and most importantly, an affirmation that they will not use or disclose any identifying information that may be derived from examination of the assessment materials. Researchers who violate the confidentiality law are subject to the same criminal penalties—fines and prison terms—as government and contractor employees.

Related Questions:

Question 3: *Can the public examine the NAEP questions and find out how well individual students performed on the NAEP assessment?*

Question 4: *Why are NAEP questions kept confidential?*

question 16

Bibliography

Allen, N.L., Donoghue, J.R., and Schoeps, T.L. (2001). *The NAEP 1998 Technical Report* (NCES 200 1–509). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Bloxom, B.P.J., Nicewander, W.A., and Yan, D. (1995). Linking to a Large-Scale Assessment: An Empirical Evaluation. *Journal of Educational and Behavioral Statistics, 20*, 1–26.

Bock, D.B., and Zimowski, M.F. (1998). *Feasibility Studies of Two-Stage Testing in Large-Scale Educational Assessment: Implications for NAEP.* Commissioned by the NAEP Validity Studies Panel. Palo Alto, CA: American Institutes for Research.

Bourque, M.L., and Byrd, S. (Eds.). (2000). *Student Performance Standards on the National Assessment of Educational Progress: Affirmations and Improvements.* Washington, DC: National Assessment Governing Board.

Bourque, M.L., Campagne, A.B., and Crissman, S. (1997). *1996 Science Performance Standards: Achievement Results for the Nation and the States.* Washington, DC: National Assessment Governing Board.

Campbell, J.R., and Donahue, P.L. (1997). *Students Selecting Stories: The Effects of Choice in Reading Assessment* (NCES 97–491). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Campbell, J.R., Hombo, C.M., and Mazzeo, J. (2000). *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance* (NCES 2000–469). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Chromy, J.R. (1998). *The Effects of Finite Sampling on State Assessment Sample Requirements.* Commissioned by the NAEP Validity Studies Panel. Palo Alto, CA: American Institutes for Research.

College Board. (2002). *Mathematics Framework for the 2003 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

Council of Chief State School Officers. (1999). *Science Framework for the 1996 and 2000 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

DeVito, P.J., and Koenig, J.A. (Eds.). (2001). *NAEP Reporting Practices: Investigating District-Level and Market-Basket Reporting.* Washington, DC: Committee on NAEP Reporting Practices, Board on Testing and Assessment, National Research Council.

Dossey, J.A., Mullis, I., and Jones, C.O. (1993). *Can Students Do Mathematical Problem Solving?* Washington, DC: U.S. Department of Education.

Feuer, M.J., Holland, P.W., Green, B.F., Bertenthal, M.W., and Hemphill, F.C. (Eds.). (1998). *Uncommon Measures: Equivalance and Linkage Among Educational Tests.* Washington, DC: Committee on Equivalency and Linkage of Educational Tests, National Research Council.

Glaser, R., Linn, R., and Bohrnstedt, G.W. (1997). *Assessment in Transition: Monitoring the Nation's Educational Progress. Background Studies of the Final Report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment.* New York, NY: National Academy of Education.

Hawkins, E.F., Stancavage, F., and Dossey, J.A. (1998). *School Policies Affecting Instruction in Mathematics* (NCES 98–495). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Hedges, L.V., and Vevea, J.L. (1997). *A Study of Equating in NAEP*. Commissioned by the NAEP Validity Studies Panel. Palo Alto, CA: American Institutes for Research.

Individuals with Disabilities Education Act, Pub. L. No. 105–17, 20 U.S.C. 1400 et seq. (1997).

Jaeger, R.M. (1998). *Reporting the Results of the National Assessment of Educational Progress*. Commissioned by the NAEP Validity Studies Panel. Palo Alto, CA: American Institutes for Research.

Jakwerth, P.R., Stancavage, F.B., and Reed, E.D. (1999). *An Investigation of Why Students Do Not Respond to Questions*. Commissioned by the NAEP Validity Studies Panel. Palo Alto, CA: American Institutes for Research.

Johnson, E.G., and Owen, E. (1998). *Linking the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS): A Technical Report* (NCES 98–499). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Johnson, E.G., Siegendorf, A., and Phillips, G.W. (1998). *Linking the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS): Eighth-Grade Results* (NCES 98–500). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Lindquist, M.M., Dossey, J.A., and Mullis, I. (1995). *Reaching Standards: A Progress Report on Mathematics. A Policy Information Perspective*. Princeton, NJ: Educational Testing Service.

Linn, R.L. (1993). Linking Results of Distinct Assessments. *Applied Measurement in Education, 6*, 83–1 02.

Linn, R.L., and Kiplinger, V.L. (1993). *Linking Statewide Tests to the National Assessment of Educational Progress: Stability of Results*. Boulder, CO: Center for Research on Evaluation, Standards, and Student Testing.

Linn, R.L., Koretz, D., and Baker, E.L. (1996). *Assessing the Validity of the National Assessment of Educational Progress: NAEP Technical Review Panel White Paper*. Washington, DC: U.S. Department of Education.

Loomis, S.C., and Bourque, M.L. (Eds.). (2001). *National Assessment of Educational Progress Achievement Levels, 1992–1998 for Civics*. Washington, DC: National Assessment Governing Board.

Loomis, S.C., and Bourque, M.L. (Eds.). (2001). *National Assessment of Educational Progress Achievement Levels, 1992–1998 for Geography*. Washington, DC: National Assessment Governing Board.

Loomis, S.C., and Bourque, M.L. (Eds.). (2001). *National Assessment of Educational Progress Achievement Levels, 1992–1998 for Mathematics*. Washington, DC: National Assessment Governing Board.

Loomis, S.C., and Bourque, M.L. (Eds.). (2001). *National Assessment of Educational Progress Achievement Levels, 1992–1998 for Reading*. Washington, DC: National Assessment Governing Board.

Loomis, S.C., and Bourque, M.L. (Eds.). (2001). *National Assessment of Educational Progress Achievement Levels, 1992–1998 for Science*. Washington, DC: National Assessment Governing Board.

bibliography

Loomis, S.C., and Bourque, M.L. (Eds.). (2001). *National Assessment of Educational Progress Achievement Levels, 1992–1998 for U.S. History*. Washington, DC: National Assessment Governing Board.

Loomis, S.C., and Bourque, M.L. (Eds.). (2001). *National Assessment of Educational Progress Achievement Levels, 1992–1998 for Writing*. Washington, DC: National Assessment Governing Board.

Mazzeo, J., Carlson, J.E., Voelkl, K.E., and Lutkus, A.D. (1999). *Increasing the Participation of Special Needs Students in NAEP: A Report on 1996 NAEP Research Activities* (NCES 2000–473). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

McLaughlin, D.H., Bandeira de Mello, V., Blankenship, C., Chaney, K., Esra, P., Hikawa, H. Rojaus, D., Williams, P., and Wolman, M (2007). *Comparison between NAEP and state mathematics assessment results: 2003* (NCES 2007-471). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC

McLaughlin, D.H., Bandeira de Mello, V., Blankenship, C., Chaney, K., Esra, P., Hikawa, H. Rojaus, D., Williams, P., and Wolman, M (2007). *Comparison between NAEP and state reading assessment results: 2003* (NCES 2007-472). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC

McLaughlin, D.H. (1998a). *Study of the Linkages of 1996 NAEP and State Mathematics Assessments in Four States*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

McLaughlin, D.H. (1998b). *Linking State Assessments of NAEP: A Study of the 1996 Mathematics Assessment*. Paper presented at the meeting of the American Educational Research Association, San Diego, CA.

Mislevy, R.J. (1992). *Linking Educational Assessments*. Princeton, NJ: Educational Testing Service.

Mitchell, J.H., Hawkins, E.F., Jakwerth, P., Stancavage, F.B., and Dossey, J.A. (1999). *Student Work and Teacher Practices in Mathematics* (NCES 1999–453). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Mullis, I.V.S. (1997). *Optimizing State NAEP: Issues and Possible Improvements*. Commissioned by the NAEP Validity Studies Panel. Palo Alto, CA: American Institutes for Research.

Mullis, I.V.S., Jenkins, F.L., and Johnson, G. (1994). *Effective Schools in Mathematics: Perspectives From the NAEP 1992 Assessment* (NCES 94–701). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

National Assessment Governing Board. (2002). *Reading Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: Author.

National Center for Education Statistics. *Mapping 2005 State Proficiency Standards Onto the NAEP Scales* (NCES 2007–482). U.S. Department of Education, National Center for Education Statistics, Washington, D.C.: U.S. Government Printing Office.

bibliography

O'Sullivan, C.Y., Weiss, A.R., and Askew, J.M. (1998). *Students Learning Science: A Report on Policies and Practices in U.S. Schools* (NCES 98–493). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Pashley, P.J., and Phillips, G.W. (1993). *Toward World-Class Standards: A Research Study Linking International and National Assessments*. Princeton, NJ: Educational Testing Service.

Pearson, D.P., and Garavaglia, D.R. (1997). *Improving the Information Value of Performance Items in Large-Scale Assessments*. Commissioned by the NAEP Validity Studies Panel. Palo Alto, CA: American Institutes for Research.

Pellegrino, J.W., Jones, L.R., and Mitchell, K.J. (Eds). (1999). *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress*. Washington, DC: National Academy Press.

Phillips, G. W. (2007). *Chance Favors the Prepared Mind: Mathematics and Science Indicators for Comparing States and Nations*. Washington, D.C. American Institutes for Research.

Shepard, L.A. (1997). Measuring Achievement: What Does it Mean to Test for Robust Understanding? *Wiliam H. Angoff Memorial Lecture Series*. Princeton, NJ: Educational Testing Service.

Shepard, L.A., Glaser, R., Linn, R., and Bohrnstedt, G.W. (1993). Setting Performance Standards for Student Achievement. *Report of the NAEP Panel on the Evaluation of the NAEP Trial State Assessment: An Evaluation of the 1992 Achievement Levels*. New York, NY: National Academy of Education.

Vinovskis, M.A. (1998). *Overseeing the Nation's Report Card*. Washington, DC: National Assessment Governing Board.

Williams, V.S.L., Billeaud, K., Davis, L.A., Thissen, D., and Sanford, E. (1995). Projecting to the NAEP Scale: Results from the North Carolina End-of-Grade Testing Program. *Journal of Educational Measurement, (35)* 4, 277–96.

bibliography

glossary

achievement levels. Performance standards, set by the National Assessment Governing Board, that provide a context for interpreting student performance on NAEP, based on recommendations from panels of educators and members of the public.

adequate yearly progress standard. The measure by which schools, districts, and states are held accountable for student performance under Title I of the No Child Left Behind Act of 2001. A state definition of AYP is based on the statewide accountability system, student achievement measurements such as test scores and graduation rates, and statewide academic assessments at the elementary and secondary levels.

assessment session. The period of time during which a test booklet is administered to students.

background questionnaires. The instruments used to collect information about student demographics and educational experiences.

bias. In a test, a systematic error in a test score. In a linkage, a systematic difference in linked values for different subgroups of test takers. Bias usually favors one group of test takers over another.

BIB (Balanced Incomplete Block) spiraling. A complex variant of matrix sampling in which items are administered so that each pair of question blocks is dispensed to a nationally representative sample of respondents.

block. A group of assessment questions created by dividing the question pool for an age or grade into subsets. Blocks are used in the implementation of the BIB spiral sample design.

booklet. The portion of the assessment instrument given to individual students created by combining blocks of assessment questions.

calibrate. To estimate the parameters of a set of questions using responses of a sample of examinees.

calibration sets. Sets of approximately 10 to 20 papers chosen by the trainer (from the training trend set or current-year responses) that serve as tools to prevent scorer drift from the standards exemplified in the scoring guide and anchor and practice papers.

composite scale. An overall subject-area scale based on the weighted average of the scales that are used to summarize performance on the primary dimensions of the curricular framework for the subject-area assessment. For example, the mathematics composite scale is a weighted average of five content-area scales: number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions. These five scales correspond to the five content-area dimensions of the NAEP mathematics framework.

constructed-response question. A non-multiple-choice question or exercise that requires some type of written or oral response.

content domain. A content domain is a set of skills and/or knowledge that is uniquely distinguished from other sets. An example of a content domain is algebra, which is distinguished from other content domains, such as geometry.

Differential Item Functioning (DIF). An item exhibits differential item functioning if the probability of doing well on the item depends on group membership, even after controlling for overall performance.

education agency. An organization involved with education administration. This could be a Local Education Agency (LEA) such as a school district, or a State Education Agency (SEA) such as a state's Department of Education.

equipercentile equating. A type of nonlinear equating in which the entire score distribution of one test is ad-

justed to match the entire score distribution of the other for a given population. Scores at the same percentile on two different test forms are made equivalent.

excluded students. Sampled students determined by the local school (using the student's Individualized Education Program (IEP) and explicit NAEP criteria) to be unable to participate meaningfully in the assessment because of a disability or because they are English language learners.

field test. Items in NAEP mathematics and reading assessments at grades 4 and 8 go through two levels of pretesting: a pilot test and a field test. A field test is the second stage of pretesting and is given 1 year prior to the full scale NAEP assessment. At a field test, the student assessment instrument for the following year is finalized. The instrument is administered to a nationally representative sample of students, and Item Response Theory (IRT) scaling decisions are made using the response data. NOTE: Previously, the term "field test" was used to refer to the first stage of item tryout in all NAEP subject-area assessments. However, beginning with the 2003 assessments, the term applies only to reading and mathematics. The stage of testing formerly referred to as a field test, starting in 2003 and in all future assessments, will be referred to as the "pilot test." All items in NAEP assessments are pilot tested, but only reading and mathematics are field tested.

framework. The blueprint, developed by the National Assessment Governing Board, that guides the development of the NAEP assessment instrument and determines the content to be assessed.

group effect. The difference between the mean for a specific group and the mean for the nation.

image-based scoring. A system used by NAEP scorers in which student response booklets are scanned, constructed responses are digitized, and the images are stored for presentation on a scorer's computer screen.

Individualized Education Plan (IEP). A program generally created for each public school student who receives special education and related services. It specifies any accommodations needed in order for the student to participate in standardized tests such as NAEP.

Item Response Theory (IRT). Test analysis procedures that assume a mathematical model for the probability that a given examinee will respond correctly to a given exercise.

large central city. A comparison group that includes public schools located in large central cities (population of 250,000 or more) throughout the United States within metropolitan statistical areas as defined by the federal Office of Management and Budget. It is not synonymous with the term inner city.

matrix sampling. A systematic way of assigning samples of test questions to different students.

multiple-choice item. An item that consists of one or more introductory sentences followed by a list of response options that include the correct answer and several incorrect alternatives.

NAEP scales. The scales common across age or grade levels and assessment years used to report NAEP results.

nonresponse. The failure to obtain responses or measurements for all sample elements.

nonresponse bias. Occurs when the observed value deviates from the population parameter due to differences between respondents and nonrespondents. Nonresponse bias is likely to occur as a result of not obtaining 100 percent response from the selected cases.

nonsampling error. A general term applying to all sources of error, with the exception of sampling error. Includes errors from defects in the sampling frame, response or measurement errors, and mistakes in processing the data.

objective. A desirable education goal accepted by scholars in the field, educators,

glossary

glossary

and concerned laypersons and established through a consensus approach.

options. The correct and incorrect response choices included in a multiple-choice question.

oversampling. Deliberately sampling a portion of the population at a higher rate than the remainder of the population.

pilot test. A pretest of questions to obtain information regarding clarity, difficulty levels, timing, feasibility, and special administrative situations. The pilot test is performed before revising and selecting the questions to be used in the assessment.

point estimate. The use of a value of a particular sample statistic to estimate the value for a parameter of interest.

poststratification. A common technique in survey analysis for incorporating the population distribution of imporant characteristics into survey estimates. Poststratification can improve the accuracy of survey estimates both by reducing bias and by increasing precision. It also corrects for nonresponse bias.

Primary Sampling Unit (PSU). The basic geographic sampling unit for NAEP. A PSU can be either a single county or a set of contiguous counties.

probability sample. A sample in which every element of the population has a known, nonzero probability of being selected.

psychometric. The field of study concerned with the theory and technique of educational and psychological measurement, which includes the measurement of knowledge, abilities, attitudes, and personality traits.

random variable. A variable that takes on any value of a specified set with a particular probability.

region. A NAEP reporting group. One of four geographic areas defined by the Office of Business Economics in the U.S. Department of Commerce, used in gathering and reporting data. These regions are the Northeast, South, Midwest, and West.

regression analysis. A statistical procedure for determining the relationship between a set of outcomes and a set of predictors. In the most common case, a single outcome (e.g., student reading proficiency) is predicted by a set of individuals' characteristics (e.g., student age, gender, and socioeconomic status).

respondent. A person who is eligible for NAEP, is in the sample, and responds by completing one or more questions in an assessment booklet.

SD/ELL student questionnaire. An instrument completed by local school staff for each student with a disability (SD) or who is an English language learner (ELL) and was selected to participate, regardless of whether or not the student was included in the assessment.

sample. A portion of a population, or a subset from a set of units, that is selected by some probability mechanism for the purpose of investigating the properties of the population. NAEP does not assess an entire population but rather selects a representative sample from the group to answer assessment questions.

sampling error. The error in survey estimates that occurs because only a sample of the population is observed. Measured by sampling standard error.

sampling frame. The list of sampling units from which the sample is selected.

sampling weight. A multiplicative factor equal to the reciprocal of the probability of a respondent being selected for assessment, with adjustment for nonresponse and, perhaps, poststratification. The sum of the weights provides an estimate of the number of persons in the population represented by respondents in the sample.

school questionnaire. A questionnaire completed for each sampled school by the principal or other official. It is used to gather information concerning school administration, staffing patterns, curriculum, and student services.

secondary-use data files.
Computer files containing respondent-level subject-area, demographic, and background data. They are available for use by researchers wishing to perform analyses of NAEP data.

selection probability. The chance a particular sampling unit has of being selected in the sample.

session. A group of students reporting for the administration of an assessment. Most schools conduct only one session, but some large schools conduct as many as 10 or more.

simple random sample. The process for selecting n sampling units from a population of N sampling units, so that each sampling unit has an equal chance of being in the sample and every combination of n sampling units has the same chance of being in the sample chosen.

specifications. The mix of item formats, the item distribution for subject-specific content areas, and the conditions under which items are presented to students (e.g., use of manipulatives, use of calculators, and length of time to complete tasks), as presented by the National Assessment Governing Board in the assessment frameworks.

split-sample design. In a split-sample design, the sample of students or schools is split into two equivalent samples that can be compared against each other. The two samples each can be assessed under different procedures and a comparison can be made. An example is the use of assessment accommodations for students with disabilities, where one sample is allowed accommodations and the other is not.

standard deviation. An index of the degree to which a set of data values is concentrated about its mean. Sometimes referred to as "spread." The standard deviation measures the variability in a distribution of quantities. Distributions with relatively small standard deviations are relatively concentrated; larger standard deviations signify greater variability. In common distributions, like the mathematically defined "normal distribution," roughly 67% of the quantities are within 1 standard deviation from the mean; about 95% are within 2 standard deviations; nearly all are within 3 standard deviations.

standard error. A measure of sampling variability and measurement error for a statistic. Standard errors in NAEP reflect NAEP's complex sample design. Standard errors may also include a component due to the error of measurement of individual scores estimated using plausible values.

standing committee. A group of teachers and education administrators convened to serve an advisory role during item development in each subject area.

statistical significance. The statistical significance of a result is the probability that the observed relationship (e.g., between variables) or a difference (e.g., between means) in a sample occurred by pure chance, and that in the population from which the sample was drawn, no such relationship or differences exist.

stratification. The division of a population into parts, or strata.

student group. Groups within the national population for which NAEP data are reported (for example, gender, race/ethnicity, grade, age, level of parental education, region, and type of location).

student ID number. A unique identification number assigned to each respondent to preserve his or her anonymity. NAEP does not record the names of any respondents.

subject area. One of the areas assessed by NAEP, including art, civics, geography, mathematics, music, reading, science, U.S. history, and writing.

systematic sample (systematic random sample). A sample selected by a systematic method (for example, units selected from a list at equally spaced intervals).

glossary

glossary

teacher questionnaire. A questionnaire completed by selected teachers of sampled students. It is used to gather information concerning teachers' educational background and experience, professional development, and classroom practices.

Title I. The primary purpose of the Title I program of the Elementary and Secondary Education Act (ESEA) is to ensure equal educational opportunity for all children regardless of socioeconomic background and to close the achievement gap between poor and affluent children, by providing resources to schools attended by disadvantaged students.

trimming. A process by which extreme weights are reduced (trimmed) to diminish the effect of extreme values on estimates and estimated variances.

variance. The average of the squared deviations of a random variable from the expected value of the variable. The variance of an estimate is the squared standard error of the estimate.

Index

Schedule of Assessments 2005 to 2017 (as of March 2009)

Year	National Assessment	State Assessment
2005	Reading **MATHEMATICS** Science High School Transcript Study	Reading (4, 8) **MATH (4, 8)** Science (4, 8)
2006	U.S. History Civics **ECONOMICS (12)**	
2007	Reading (4, 8) Mathematics (4, 8) Writing (8, 12)	Reading (4, 8) Math (4, 8) Writing (8)
2008	Arts (8) Long-term trend	
2009	**READING** Mathematics* **SCIENCE** High School Transcript Study	**READING (4,8,12)** Math (4, 8, **12**) **SCIENCE (4, 8)**
2010	U.S. History Civics Geography	
2011	Reading (4, 8) Mathematics (4, 8) **WRITING**	Reading (4, 8) Math (4, 8) **WRITING (4, 8)**
2012	Economics (12) **PROBE: TECHNOLOGICAL LITERACY [special study]** Long-term trend	
2013	Reading Mathematics Science High School Transcript Study	Reading (4, 8) Math (4, 8) Science (4, 8)
2014	**U.S. HISTORY** **CIVICS** Geography	
2015	Reading (4, 8) Mathematics (4, 8) Writing	Reading (4, 8) Math (4, 8) Writing (4, 8)
2016	Arts (8) Long-term trend	
2017	Reading Mathematics Science High School Transcript Study	Reading (4, 8) Math (4, 8) Science (4, 8)

*New framework for grade 12 only.

NOTES:

(1) Grades tested are 4, 8, and 12 unless otherwise indicated, except that long-term trend assessments sample students at ages 9, 13, and 17 and are conducted in reading and mathematics.

(2) Subjects in **BOLD ALL CAPS** indicate the year in which a new framework is implemented or assessment year for which the Board will decide whether a new or updated framework is needed.

For a complete list of subjects assessed prior to 2000, consult the NAEP website at http://nces.ed.gov/nationsreportcard/about/assessmentsched.asp.

schedule of assessments

Appendix B

Home

Reviewing your school's assessment systems

Schools are no longer required to use National Standards (archived) for assessing and reporting on progress and achievement in literacy and numeracy for years 1–8, and the National Administration Guidelines (NAGs) have been revised to reflect this.

The NAGs now state that schools must collect, analyse, and report on good quality assessment information. This is defined as assessment information that "draws on a range of evidence to evaluate the progress and achievement of students and build a comprehensive picture of student learning across the curriculum".

Furthermore, there is the requirement in NAG 2 to:

(c) On the basis of good quality assessment information, **report to students and their parents** on progress and achievement of individual students:

- in plain language, in writing, and at least twice a year; and
- **across The National Curriculum**, as expressed in *The New Zealand Curriculum* or Te Marautanga o Aotearoa, including in mathematics and literacy, and/or te reo matatini and pāngarau.

There is room for some interpretation as to what will adequately give effect to reporting to parents and their families "across the The National Curriculum", so schools do have discretion to design both their approach to the curriculum and to assessing and reporting on student progress and achievement, so that both recognise the aspirations and desires of the local community.

Schools now have more flexibility about how they assess, what assessment information they collect and analyse, and how they use it. If you choose to review your approach, it is still important to ensure that any changes retain at least the same rigour and educative purpose as before.

These comments from the May 2018 ERO report on assessment in primary schools are pertinent:

> "If we are to improve the success of all learners and enable them to achieve the outcomes inherent in the New Zealand Curriculum, we also need purposeful leadership focused on improving students' learning, and better use of information to make appropriate decisions for, about and with students. Assessment literacy, use of assessment data, school leadership, boards of trustee capacity to enquire into school performance data and student progress, and school planning and reporting remain as key challenges to lifting student achievement and thereby system performance."
>
> "Much assessment in primary schools has recently focused on reading, writing and mathematics, to help children develop the literacy and numeracy skills needed to fully engage with the whole curriculum. Knowledge and confidence in these areas is crucial. However, some schools are going much further usefully identifying samples of work that demonstrate students' confidence with Key Competencies from *The New Zealand Curriculum*. It is now timely to consider extending assessment practices, to determine how well students are progressing in applying their skills to meaningful tasks from other curriculum areas and key competencies."
>
> "As a system we need urgent agreement in respect to how we will measure, monitor and report on student progress across the curriculum."
>
> "Equally the sharing of information between schools is limited. There is a need for nation wide agreement on the appropriate form and content of assessment information which accompanies students as they progress through the system or move from school to school."

Reviewing an assessment approach can be complex and time consuming. We are presenting a three step approach that we think will make the process clear to all and as time-efficient as possible. Good luck.

Consider:

The intertwined nature of curriculum and assessment.

- What is your interpretation of the curriculum, designed to best meet the needs of your students and community?
- How do you ensure that every student is well supported to learn all that the curriculum offers?
- Do you need to review these?

If you do, then how you assess will also need to be reviewed.

Re-refresh your familiarity with these documents:

- Purposes of Assessment (DANZ report)
- Changes to National Administration Guidelines, particularly NAG 2

- [Principles of assessment](#) (NZCurriculum)
- Users of assessment information ([New Zealand Curriculum, p.40](#))

Once you have finished implementing your review findings, you will want your system to align with the expectations described in these documents.

Review:

- **Understanding progress** – do teachers have a common understanding about how student learning builds and grows? Are big ideas and competencies clear at each level? Is there a shared language of learning across the school? Can teachers locate a student's performance along a continuum of curriculum levels?
- **Current pedagogy** – how well are student agency and assessment capability being grown across the school? How well does assessment for learning underpin teaching and learning?
- **Current assessment activities and tools** – what works, what doesn't, what needs modifying?
- **Quality assessment processes that ensure dependable information** – assessment administration, moderation, data entry, data analysis. What works, what doesn't, what needs modifying?
- **Current processes for sharing information** – is achievement information shared with those who need it, what processes work well, what needs modifying?

Implement as appropriate:

- school-wide clarification of what progress looks like, taking into account the 'front end' of the curriculum as well as curriculum achievement objectives
- processes to improve effective assessment for learning in classrooms
- improved streamlined, skilled use of assessment tools
- new assessment tools and approaches
- use of assessment information for improvement – including sharing information with parents/whānau and board and community. Our [Why report to parents and whānau page](#) shares ways to share information for learning, rather than for compliance.

Never believe that any review, no matter how thorough, is the final product and that it is finished. The curriculum keeps evolving, and so too must how we represent and monitor progress in the learning of it.

Key ideas for guidance:

- Keep the [principles of assessment for learning](#) in mind and check the coherence of your emerging system against them rigorously.
- Keep the stakeholders for assessment information clearly to the fore. As you design your system, make sure that you can meet the legitimate needs of all of them:
    - students
    - teachers
    - school leaders and board of trustees
    - parents and whānau
    - school or kāhui ako.
- Think carefully about your purposes for assessment. Assess only when the information will be used to improve teaching and learning. **Do not over assess.**
- Think carefully about your processes for recording, storing and sharing assessment information. These should be easily updated and accessed via your SMS or other technology.
- Keep overall judgments of progress and achievement as a central part of the system at all levels. Remember that it is the big ideas, the concepts, that we want all learners to take hold of, as well as the techniques, skills and strategies.
- Remember that moderation, using agreed sources of evidence, is vital in order to keep everyone assured of the dependability of overall judgments about where the learner is at.
- Remember that your assessment systems must be able to measure both progress and achievement.
- Don't throw the baby out with the bath water. Whatever you have been doing that you cherished while [National Standards (archived)](#) were with us – keep. Change the bits that don't work so well. Discard the bits you don't like and that obfuscate learning.

Developing curriculum

The updated NAG requirements might cause you to rethink your curriculum plan and how you will think about representing progress and achievement. If your school has a Graduate Profile, this is good time to be thinking about what that profile looks like, not only at graduation, but also at key year levels or transition points: for example, from year 3 to 4, year 6 to 7, and year 8 to 9.

Ideas about [Key competencies](#) have been evolving. Think about the big capabilities you want for your students at those transition points through the school that will result in graduates who have the capabilities described in your profile. Some of these [school stories](#) share their early thinking about how they are going about this. Rather than trying to develop learning progressions for all of the subject areas, which is too big a job for individual

schools, this is a good time to be thinking carefully about the overall concepts.

The [Cross-sector forum toolkit](#) developed for Communities of learning | Kāhui Ako contains a set of coherent pathways shaped around four capabilities (big ideas), described for several key transition points. You can read about it in this document:

[Capabilities across the curriculum](#) (PDF 267 KB)

This might be a useful place to start if you don't already have a graduate profile or similar.

Making teacher judgments on progress and achievement

No single source of information can accurately summarise a student's progress or achievement. A range of approaches is necessary in order to compile a comprehensive picture of the areas of progress, areas requiring attention, and what a student's unique progress looks like. Using a range of approaches allows the student to participate throughout the assessment process, building their assessment capability.

The three elements that comprise teacher judgment of student work, according to Sadler (1988) are:

- teachers attending to the learning students produce
- appraising this work against a reference framework
- making an explicit response such as feedback or judgment on the learner's work.

Teacher judgment for sharing information on progress and achievement for reporting purposes has an added complexity because the judgment is made on a range of evidence. Teachers draw on and apply the evidence gathered up to a particular point in time, in order to make a judgment about a student's progress and achievement.

The use of a range of evidence builds dependability in decisions. A teacher judgment can be made when the teacher reviews all of the evidence in relation to the benchmarks decided upon by the school, be they curriculum levels or progressions of learning.

Evidence can be gathered in a range of ways, by:

- **observing** the process a student uses to complete a learning task, such as:
  - focused classroom observation
  - student work books
  - rich tasks
  - running records
  - self and peer assessment
- **conversing** with the student to find out what they know, understand and can do.
  - conferencing
  - interviewing
  - questioning
  - explaining
  - discussing
- **gathering** results from informal and formal assessments, including standardised tools.
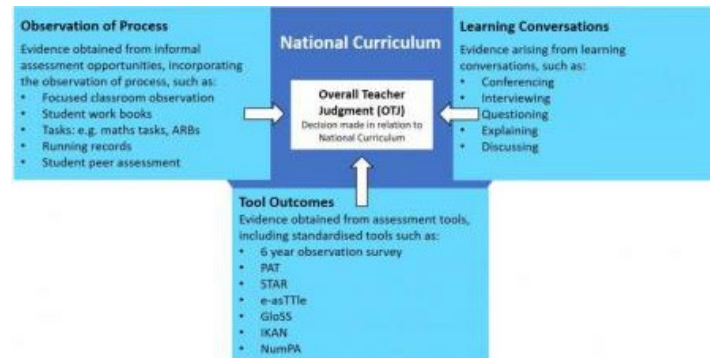
Select the image to view at full size.

School example

This article, Using a range of information – Reading example shows how evidence is collected from multiple sources in reading. You can see clearly from the example that students are fully involved in the learning and assessment process.

Increasing the dependability of teacher judgments

Increasing the dependability of evidence from all sources is fundamental to reaching a valid and defensible judgment on student progress and achievement. Students' performance in assessments will vary from day to day depending on:

- the nature of the assessment task
- the conditions in which the assessment is undertaken
- the purpose of the assessment
- the student's preparation
- the student's engagement and motivation.

When teachers experience some degree of inconsistency with assessment information, they should inquire into this further. If the inconsistency cannot be explained by normal variation in students' performance, then there may be a need to collect further information in order to reach robust judgments.

Moderation

Moderation can help to improve the dependability of teacher judgments, and of the evidence that informs and supports them. Teachers should moderate

their assessments and their overall judgments of progress and achievement in relation to school benchmarks. This process is discussed in more detail in the Moderation section of this site.

Student participation

Students should actively participate throughout the assessment process. This is a vitally important characteristic of effective assessment for learning.

It benefits the students' assessment capability by clarifying what they know, understand and can do, and what they need to learn next. All students can participate in the assessment process to some degree. As their assessment capability grows and develops, they can become more and more actively involved.

Including the students in the judgment-making process will also give them confidence to talk about their progress and achievement with their parents, family and whānau.

Involving students – an example illustrates how, through conversation, year 5–6 students become clearer about their learning, progress, and achievement.

Summary

- Making judgments about progress and achievement involves both student and teacher.
- There is a need for information from a range of assessment approaches so that decisions are dependable.
- Moderation improves both the dependability of teacher judgments and the evidence that supports them.
- Teacher curriculum and pedagogical content knowledge is essential for making dependable judgments.
- How might I involve students in the process of making judgments on their progress and achievement in my classroom?
- How much information might I need to make a judgment?
- How might I support the judgments I make and how dependable are they?
- What do I need to do to increase the dependability of my judgments?
- How does our school manage the moderation of assessments and judgments?

Questions for discussion

- How might I involve students in the process of making judgments on their progress and achievement in my classroom?
- How much information might I need to make a judgment?
- How might I support the judgments I make and how dependable are they?
- What do I need to do to increase the dependability of my judgments?
- How does our school manage the moderation of assessments and judgments?

Measuring progress across the curriculum

The New Zealand Curriculum sets out broad achievement objectives for each of the learning areas and for each curriculum level. But it is a framework rather than a detailed plan. This means that while every school curriculum must be clearly aligned with the intent of the New Zealand curriculum, schools have considerable flexibility when determining the detail of teaching, learning and assessment in their context. In doing this, they can draw on a wide range of ideas, resources and models.

⊖ Resources available across the curriculum

For teachers in years 1 to 8, measuring progress across the curriculum is more problematic than in literacy and mathematics. Some learning areas are well supported by learning progressions, others less so. For some learning areas, schools will have to make decisions based on Achievement Objectives and available resources, and develop assessments accordingly.

This is an opportunity for schools to work collaboratively to develop their learning objectives and progress indicators. The Kāhui Ako is an ideal vehicle for this.

Schools will need to think carefully about which areas of the curriculum and at what levels they should assess. For example, curriculum and assessment emphases at year 1 and year 8 will probably be different. Emphases will be different for different contexts and schools.

It's important that data about student progress and achievement can be collated and analysed at a class and cohort level, to allow for the identification of students who are not making sufficient progress or who have special needs. This data may also identify aspects of the curriculum that require attention: for example, extra professional learning and development in a particular curriculum area. You can find lots of information on gathering and analysing data in the Using evidence for learning section of this website.

Below is a collection of the resources available that schools or Kāhui Ako could use.

Science

Planning, teaching and possible assessment resources available:

- Science Online TKI
- Science Learning Hub
- NZ Curriculum Exemplars – although these **exemplars** are resources archived on the old TKI site, and relate to the previous curriculum, they are valuable in assisting schools to define progressions of learning and fine-grained learning objectives, against which they can assess progress and achievement. Have a look at the matrices of progression from which you could develop your own, in conjunction with the AO's from the new curriculum.

Assessment tools:

- Assessment Resource Banks
- Junior Science: Thinking with Evidence
- Science: Thinking with Evidence

Read this school story from Leeston School. It tells how the school uses **rubrics** to assess aspects of the Nature of Science (NoS).

This short paper from Rose Hipkins of NZCER highlights the difficulties of assessing science against the AOs in the curriculum – 'Problems with reporting science in curriculum levels'.

....

Social studies

Planning, teaching and possible assessment resources available:

- Social Sciences Online
- NZ Curriculum Exemplars – The Social Sciences Online website states that *Resources and information relating to the previous (1997) curriculum document may continue to be used as supporting documents when planning teaching and learning programmes.*
  - Although the exemplars are resources archived on the old TKI site, and relate to the previous curriculum, they are valuable in assisting schools

to define progressions of learning and fine-grained learning objectives. Have a look at the Social Studies matrices of progression from which you could develop your own, in conjunction with the AOs from the new curriculum.

- There are some assessment resources for years 4 to 8 available from the National Education Monitoring Project, since superseded by the NZMSSA. It is hoped that resources from the new project will be made available to schools soon. Although NMSSA tasks have not yet been made available for teachers, the reports from the monitoring study make interesting assessment reading.

Activities to support rich learning: Assessing progression in social studies

This paper by Samantha Sasse and Bronwyn Wood, published on ResearchGate, explores a new format for assessing primary social studies based on a review of 220 videos from NMSSA. The paper has since been published in SET 3, 2017 by NZCER.

Technology

Planning, teaching and possible assessment resources available:

Technology Online provides a range of resources to assist teachers of technology:

- Indicators of progression have been developed in technology to help teachers mediate the achievement objectives into specific Learning Outcomes. The indicators can be used to plan learning experiences, aid in diagnostic assessment, and support formative interactions within the classroom to help scaffold student learning. They can also support summative assessment for reporting purposes. The indicators are indicative of the level expected by the achievement objective.
- Progress outcomes, exemplars and snapshots have been developed for the new Digital Technology section of the technology curriculum.
  - Computational thinking for digital technologies
  - Designing and developing digital outcomes
  - Technology Online webinar: Getting started with the learning progressions for digital technologies in primary

In this video, Cheryl Pym explains how to track student progress in technology.

The Arts

Planning, teaching and possible assessment resources available:

- NZ Curriculum Exemplars: Although the National exemplars are archived on the old TKI site, and relate to the previous curriculum, they are still valuable in assisting schools to define progressions of learning and fine-grained learning objectives. Have a look at the matrices of progression from which you could develop your own, in conjunction with the AOs from the new curriculum.
- Arts Online TKI provides resources which can help teachers put together programmes of learning and assessment in the arts.

Check out the school story page to read about how Owhango School is bringing the arts back into the curriculum.

Health and Physical Education

Planning, teaching and possible assessment resources available:

- NZ Curriculum Exemplars: Although the National Exemplars are archived on the old TKI site and relate to the previous curriculum, they are still valuable in assisting schools to define progressions of learning and fine-grained learning objectives, against which they can assess progress and achievement. Have a look at the matrices of progression from which you could develop your own, in conjunction with the AOs from the new curriculum.
- Health and Physical Education TKI provides resources which could help schools plan their teaching, learning and assessment programme.

Consult the Assessment tool selector and the Assessment resources maps for all available assessment tools in the various subject areas.

⊕ Resources available in literacy and mathematics

Rather than assessing key competencies as stand-alone capabilities, schools will want to investigate ways in which they can weave them into the learning areas so that they become an intrinsic part of the learning and assessment. In this article, *Weaving a Coherent Curriculum*, Rose Hipkins explores how to build capability by weaving curriculum design with key competencies.

Appendix C

Scottish National Standardised Assessments

What are the Scottish National Standardised Assessments (SNSA)?

School children in P1, P4, P7 and S3 complete online standardised assessments in literacy and numeracy as part of everyday learning and teaching.

The assessments help to identify children's progress, providing diagnostic information to support teachers' professional judgement.

The assessments are as inclusive as possible to accommodate the needs of children and young people who require additional support.

There is no pass or fail and children and young people do not have to revise or prepare for these assessments.

Why have these assessments been introduced in Scottish schools?

The SNSA has been introduced as part of the National Improvement Framework for Scottish Education (http://www.gov.scot/Topics/Education/Schools/NationalImproveme to provide teachers with objective, comparable information about progress, which will help improve outcomes for all children and young people.

Assessment is a central part of everyday learning and teaching for every child and young person and these assessments have been specifically designed to reflect the way we deliver education in Scotland.

Ongoing and periodic assessments are, and will continue to be, the main basis of teachers' professional judgement. Teachers should continue to draw on all of the assessment information available to them when considering children's progress and when planning next steps in learning.

The SNSA assessments have been developed to align with Curriculum for Excellence.

What **you** need to know

An overview of assessments

SNSA provides Scottish teachers with diagnostic information on aspects of Reading, Writing and Numeracy at P1, P4, P7 and S3 stages. The assessments are marked and scored automatically giving teachers immediate feedback to help children progress through their learning.

Once the assessments have been completed, a diagnostic report is available for teachers based on either an individual or groups of learners. The reports focus on establishing where learners are in their progress in literacy and numeracy, identifying strengths as well as areas that require further support. The information provided by the assessments helps teachers to assess children's progress and to plan next steps in learning.

For more information please visit the assessments and reports page (/more-about-assessments-and-reports).

Do you need more information?

Please visit the Question and answers (questions-and-answers) section or scroll down for some further reading.

If your questions are not answered there, feel free to contact the Service Desk (/contact) who will be happy to help.

Further reading

Explore further reading resources that provide the wider context for the Scottish National Standardised Assessments as well as advice on self-evaluation and improvement.

Home (/) // **Teachers**

Information for teachers

The Scottish Government commissioned SNSA as part of the National Improvement Framework. The SNSA was designed and developed specifically for the Scottish curriculum and all questions are reviewed and approved by panels of Education Scotland and Scottish Government staff. Each question is mapped to an organiser and sub-category in the Curriculum for Excellence Literacy or Numeracy benchmarks, at a relevant stage for the year group.

The SNSA complement the assessment information gathered from everyday learning situations, and through observation and interaction with children and young people. The SNSA will:

Jump to ...

- What you need to know as a teacher
- Training information

- provide diagnostic reports detailing areas where children and young people have shown particular success or where they require further development
- help you to make decisions about the next steps in learning, both for individual children and in terms of the particular approaches used in the classroom
- provide you with additional information to consider when making a professional judgement on a child's progress in achieving the relevant Curriculum for Excellence level.

What you need to know as a teacher

What is being assessed?

Scottish National Standardised Assessments focus on aspects of reading, writing and numeracy.

Who is being assessed?

Children and young people in P1, P4, P7 and S3.

What do the assessments comprise?

Children are presented with a range of questions, often with illustrations or images. As the system is adaptive, the questions change depending on how well a child is doing. This means that each child follows their own pathway through the assessments and avoids them having to face lots of questions that are too easy, or too difficult. If a child is struggling, the questions will

When are the Scottish National Standardised Assessments going to be available?

The assessments are available from August and close at the end of the academic year.

When will children take assessments?

In consultation with the school and local authority, teachers will decide when within the school session is most appropriate time. The assessments can be carried out:

- during the school year providing diagnostic information on how individual children are progressing in aspects of literacy and numeracy, and helping planning for next steps in learning.

- at the end of the school year to help inform teachers' judgements of whether individual children have achieved the Curriculum for Excellence level relevant to their stage. The results of each child's assessment should be shared with their teacher for the following year so that future learning can be planned.

get easier, and if a child is doing well, the questions will become more challenging.

What does this mean for you?

You should approach the assessments as part of normal teaching and learning activities for children and young people. No special preparation is required. However, you will need to think about how you manage administering the assessments within your classroom. Tips and illustrations on classroom management are available within the assessment platform. It is important that you ensure children and young people understand the standardised assessments are just one part of a broader approach to assessment within Scottish schools.

How will this affect how you teach?

The assessments should not require a change to the way you teach and there should be no additional workload for you or for the children. However, you will need to think about how best to manage the assessments in your particular context and plan accordingly. The assessments are as inclusive as possible to accommodate the needs of children and young

How will children be assessed?

Children will be assessed using the SNSA online system. Assessments are automatically marked online and teachers can generate reports. The assessments provide an objective and comparable measure of children's progress in aspects of reading, writing and numeracy.

How many assessments do children and young people take in a school session?

- P1 children take two SNSA assessments: one in literacy and one in numeracy.
- P4 children take three SNSA assessments: one in reading, one in writing and one in numeracy.
- P7 children take three SNSA assessments: one in reading, one in writing and one in numeracy.
- S3 young people take three SNSA assessments: one in reading, one in writing and one in numeracy.

How long do the assessments take?

The assessments are designed to be as short as possible and are age and stage appropriate. There is no time limit. This is to ensure children do not feel unnecessary time pressure when undertaking the assessments.

people who require additional support. The diagnostic reports from the assessments should provide insight into how you might plan your future teaching to meet children's needs.

What does this mean for your children and young people?

Children and young people should not revise or do extra work for these assessments and evidence shows us that many children enjoy doing them. The assessments have been designed in line with Curriculum for Excellence so daily classroom learning is all the preparation a child or young person will need.

SNSA data will contribute to a complete and balanced picture of how children are progressing, giving you diagnostic information to help support every child's learning. The assessments should not be used in isolation and should not be viewed as a replacement for the ongoing assessment of children's progress that is central to Curriculum for Excellence.

As always, your professional judgement is key in measuring a child's progress.

P1: In most cases it will take P1 learners less than 30 minutes to complete an assessment.

P4/P7/S3: On average, allow up to 40 minutes to complete the assessment. Some learners will need more time, but many will finish in less than 30 minutes. Remember that learners can return to complete an assessment after a break.

Classroom management and IT provision should be carefully considered before administering the assessments as this can affect time taken.

Can you access the assessments through Glow?

Users with a Glow login can access the assessments directly when logged into Glow. For further information, see the SNSA login page (/access-scottish-national-standardised-assessments/).

Training information

Flexible training

A flexible training programme is available for the Scottish National Standardised Assessments. Teachers who are responsible for assessing children and young people in Primary 1, Primary 4, Primary 7 or S3, headteachers, local authority staff and Education Scotland staff receive

Online training

All online resources, which can be accessed on demand, are available from within the SNSA assessment platform. These resources can be used to refresh knowledge after attending an event or provide guidance for teachers unable to attend an event. Support Officers from SCHOLAR provide SNSA training and professional development. If your authority has

tailored training through a mixture of online and face-to-face events as agreed at local authority level.

The Glow platform can also be used to access copies of the training videos (https://glowscotland.sharepoint.com/sites/PLC/snsatraining) and getting started guide (https://glowscotland.sharepoint.com/sites/PLC/snsauserguide) allowing teachers to access material at a time and place of their choosing. Note, Glow is available to authenticated users only.

Training sessions

Training and professional development for the Scottish National Standardised Assessments are available in a variety of formats including face-to-face presentations, webinars and online resources.

There are currently five training and professional development courses available:

Course 1: 'An Introduction to SNSA' is a training course which explains how to access, assign and conduct SNSA. It is suitable for those who have yet to conduct SNSA, particularly new Account Managers and Assessment Managers.

Course 2: 'An Introduction to Analysing SNSA Data' is a training course which focusses on generating

requested an SNSA webinar, you will be able to access a recording by emailing the service desk at info@snsa.org.uk. Other training materials, such as FAQs and 'Quick Start' Guides, are available within the SNSA assessment platform. Further guidance relevant to local authorities and Education Scotland is also available A technical Service Desk is available; the contact details are provided on the SNSA platform or email info@snsa.org.uk (mailto:info@snsa.org.uk).

Quick Start eModule

SCHOLAR has created a series of short videos that act as a step-by-step guide for teachers. These videos demonstrate how to access and assign assessments to children and young people. In addition, there is guidance on how to access student login details, set up specific groups (or tags) and view practice assessments. The Quick Start training is available within the 'Help' section of the SNSA assessment platform.

Service Desk

A service desk is available to all users of the Scottish National Standardised Assessments within the assessment system. The Service Desk is staffed

and analysing SNSA reports to inform learning and teaching. The course incorporates discussion of SNSA in context and how SNSA can support teacher judgement. It is suitable for those who have conducted SNSA or participated in **Course 1: 'An Introduction to SNSA'**.

Course 3: 'An Update on SNSA Data Analysis – Using the Long Scale' is an online course which explains the long scale, introduced into SNSA at the start of the 2018 to 2019 academic year. It is suitable for those who have participated or will participate in **Course 2: 'An Introduction to Analysing SNSA Data'** or **Course 4: 'Using SNSA Data to Support Improvement'** and wish to gain a comprehensive understanding of the long scale used in SNSA.

Course 4: 'Using SNSA Data to Support Improvement' is a training course which focusses on understanding the use of SNSA data in practice and how this fits within the wider NIF assessment strategy. It is suitable for those who have participated in **Course 2: 'An Introduction to Analysing Data'** (or who have viewed the 'Interpreting Reports' videos online) and are comfortable with interpreting SNSA data.

from 0800-1800 Monday to Friday. Service Desk Operators will be able to answer your questions regarding the Scottish National Standardised Assessments.

Contact details for the Service Desk are available within the assessment system. You can also email at info@snsa.org.uk (mailto:info@snsa.org.uk).

Course 5: 'A Closer Look at Using SNSA in an ASN and EAL Context' is a training course which incorporates discussion of using SNSA in practice with children and young people with additional support needs and/or English as an additional language, including accessibility guidelines and practical advice. It is suitable for those who are responsible for conducting SNSA with children and young people with additional support needs and/or English as an additional language.

Looking for the assessment system?

Log in to the Scottish National Standardised Assessment system for a quick start guide, the staff menu, help content and contact details for the SNSA Service Desk.

SNSA login ❯ (/access-scottish-national-standardised-assessments/)

Information for parents and carers

The Scottish Government commissioned the Scottish National Standardised Assessments as part of the National Improvement Framework.

These assessments provide teachers, for the first time, objective and nationally consistent information on children's progress in aspects of literacy and numeracy, alongside a wide range of other assessment activity. Teachers can then discuss children's progress with them and their parents, taking into account the full range of assessment activity, including SNSA, to plan next steps and ensuring parents understand how best to support their child's learning at home.

For children and young people to have the best chance of reaching their potential, parents, carers, teachers and the children themselves need to understand how they are progressing and what further support they require.

Jump to ...

- What you need to know as a parent or carer
- How will the new system benefit your child?
- What does this mean for you?

What you need to know as a parent or carer

What's new?

In August 2017 Scottish Government introduced the SNSA, a single, nationally developed set of standardised assessments for aspects of literacy and numeracy, designed to align with the way we deliver education in Scotland through Curriculum for Excellence.

These assessments are expected to replace the variety of existing standardised assessments that local authorities and schools use at the moment.

In addition to SNSA, ongoing and informal assessment continues to be a central part of everyday assessment in schools.

What is being assessed?

The SNSA assess aspects of reading, writing and numeracy, providing teachers with diagnostic information on children's progress in these areas.

How many assessments do children and young people take in a school session?

- P1 children take two SNSA assessments: one in literacy and one in numeracy.
- P4 children take three SNSA assessments: one in reading, one in writing and one in numeracy.
- P7 children take three SNSA assessments: one in reading, one in writing and one in numeracy.
- S3 young people take three SNSA assessments: one in reading, one in writing and one in numeracy.

What does this mean for your child?

Children do not have to revise or prepare for assessments. The assessments are used as part of routine teaching and learning to help teachers understand how well your child is progressing and to plan next steps.

Who is being assessed?

All children in P1, P4, P7 and S3 are assessed once a year in aspects of reading, writing and numeracy.

When do the assessments take place?

The assessments are administered to each child once within the relevant school year at a time chosen by the class teachers with guidance from their schools and local authorities. They decide the most appropriate time during the school year for your child to take the standardised assessments.

What form do the assessments take?

Children complete the assessments online and they are marked automatically, giving teachers immediate feedback to help children progress in literacy and numeracy.

The assessments are adaptive, which means that if a child is experiencing difficulty, the questions will get easier, and if a child is doing well, the questions will become more challenging. In this way, the assessments establish children's capacity without them having to face lots of questions that are too easy or too hard for them to answer. There is no pass or fail.

The assessments are as short as possible and are age and stage appropriate. There is no time limit. This is to ensure children do not feel unnecessary time pressure when undertaking the assessments. The majority of P1 learners will complete each assessment in less than 30 minutes, while on average P4/P7/S3 learners complete within 40 minutes. Your child will not be expected to take assessments covering reading, writing, and numeracy in one sitting.

How are the results being used?

Scottish National Standardised Assessments data provides additional information to the teacher regarding how your child is progressing in school.

How will the new system benefit your child?

How does the SNSA benefit your child?

The online assessment system produces diagnostic information about where your child did well and where further support is required.

Your child's teacher uses this feedback to help plan next steps and provide further support as appropriate. Providing the right support at the right time helps to ensure your child can reach their potential.

What does your child need to do?

The assessments are just another aspect of daily learning. Your child does not have to revise or prepare for assessments. There is no additional workload for your child and the assessments do not distract from daily learning.

What does this mean for you?

Do you get to see the results?

As part of the normal reporting process in your school teachers use this information, alongside a wide range of other assessment information, to discuss with you how your child is progressing with their learning.

How is the information gathered used?

Teachers discuss feedback from assessments with individual children and then plan next steps in learning using all available assessment information.

The information is also used by schools to plan improvements

Who can I talk to for more information?

Your child's class teacher or headteacher should always be on hand to discuss any questions that you may have about your child's progress and the role assessment plays in this. Your Parent Council may be able to point you towards further advice or support.

If you have more general comments or questions about the way in which children's progress is being assessed as part of the National Improvement Framework, please email nationalimprovementframework@gov.scot (mailto:nationalimprovementframework@gov.scot).

in literacy and numeracy provision.

How can you help?

You have a key role to play in helping your child to progress in their learning. Talk to your child's teacher if you have any concerns about their learning or assessment. Ask for information on how you can support your child's learning at home.

For further information about how to support your child's learning, their education and school life, visit **Parentzone Scotland's website (https://education.gov.scot/parentzone/)**.

Can I withdraw my child from the SNSA?

The SNSA are part of everyday learning and teaching. The assessments provide teachers with diagnostic information to help them plan next steps in children's learning. Similar to other methods of assessment in schools, there is no legal basis for a parent to withdraw their child from the SNSA. If you have specific queries or questions about your child undertaking the assessments then you should discuss this matter with your child's school who will take the decision on whether to grant the request as they would for any other aspect of learning and teaching.

Where can I find out more?

Further information about the **National Improvement Framework (http://www.gov.scot/Topics/Education/Schools/NationalImprov** be found at the **Scottish Government's website (http://www.gov.scot)**, where you can also find **Assessing Children's Progress: A guide for parents and carers (https://beta.gov.scot/publications/assessing-childrens-progress-guide-for-parents-and-carers/National%20Improvement%20Framework%20-%20assessing%20progress%20-%20guide%20for%20parents.pdf?inline=true)**.

(/)

Information for children and young people

The Scottish National Standardised Assessments system is an online tool that is fun and easy to use. It is designed to help your teachers understand exactly what they need to do to help you make good progress with your reading, writing and numeracy skills.

If you have questions, there is a lot of information on this website and your class teacher will always be happy to talk to you about the assessments.

Jump to ...

- What you need to know as a child or young person

What you need to know as a child or young

person

Who takes the assessments?

In primary school, children in P1, P4 and P7 take the assessments. In secondary school, young people in S3 take the assessments.

What about your results?

Your teachers discuss your results with you to help you understand where you are making progress and which areas need further work. There is no pass or fail and definitely nothing to worry about!

Should you prepare for the assessments?

The assessments are part of your normal classroom work and you do not need to do any special preparation for them. The assessments are as short as possible and you are allowed to take as long as you need to complete them.

You will use a computer to take the assessments. Your teachers will be there to make sure everything runs smoothly.

Contact (/contact/)

Accessibility (/accessibility/)

Privacy policy (/privacy-policy/)

≡ Q

Home (/) // **Local authorities**

Information for local authorities

The Scottish Government developed the Scottish National Standardised Assessments as part of the National Improvement Framework.

The single, nationally consistent set of standardised assessments has been designed to align with the way we deliver education in Scotland through Curriculum for Excellence, and should replace the variety of existing standardised assessments that local authorities and schools currently use.

Jump to ...

- **What you need to know as a local authority**
- **What is the rationale for the SNSA?**
- **Are the assessments inclusive and accessible?**
- **What should you do now?**

What you need to know as a local authority

What is being assessed?

The SNSA focus on aspects of reading, writing and numeracy and are aligned to the curriculum benchmarks for literacy and numeracy.

Who is being assessed?

All children in P1, P4, P7 and S3 are assessed once a year in aspects of reading, writing and numeracy.

When will the assessments take place?

Individual teachers and schools, with guidance from their local authorities, decide the most appropriate time during the school year for children to take the standardised assessments.

What form do the assessments take?

Assessments are completed online and automatically marked by the online system, giving teachers immediate feedback to help children progress.

How will this impact children?

The assessments are delivered as part of routine learning and teaching. Children are not expected to revise or prepare for assessments. The assessments are as inclusive as possible and accommodate the needs of children and young people who require additional support. Considered classroom management will ensure a positive experience for children.

How are the results be used?

SNSA data contributes to a complete, consistent and balanced picture of how children are progressing in school. Teachers, schools, and local authorities use this information to make further improvements at an individual, classroom, school and local authority level, which in turn helps every child to succeed.

What is the rationale for the SNSA?

Why introduce a national assessment system?

Consistent, objective and comparable information is required at school, local authority and national level to show us which learning and teaching approaches are working best and where changes need to be made. That way we can learn from the best and ensure time and efforts are focused on the approaches that have the greatest impact on learning.

Scottish Government has introduced a single, nationally consistent set of standardised assessments, designed to align with the way we deliver education in Scotland through Curriculum for Excellence. These assessments are replacing the variety of existing standardised assessments that local authorities and schools currently use.

How do the assessments benefit children and teachers?

The SNSA provide diagnostic information on how all children in Scotland are progressing with aspects of literacy and numeracy. Taken together with all other ongoing assessment information, the national standardised assessment data provides a more complete, consistent and balanced picture of how children are getting on.

Alongside a range of other evidence, the SNSA informs teachers' professional judgements of Curriculum for Excellence levels. The assessments should not be used in isolation and should not be viewed as a replacement for the ongoing assessment of children's progress which is central to Curriculum for Excellence.

Scottish Government is clear that the purpose of all assessment, including national standardised assessments, is to help teachers understand how children are progressing, and to help teachers tailor aspects of future learning accordingly. Assessment data

should be used to improve educational outcomes for every child and young person.

Are the assessments inclusive and accessible?

Are the assessments appropriate for children with ASN?

The SNSA are designed to be as inclusive as possible to accommodate the needs of children and young people who require additional support. Whatever support a child receives in the classroom should be available for the assessments.

An Accessibility Advisory Group is in place for the SNSA and provides valuable assistance in ensuring the system is accessible to most children. Their recommendations are being incorporated into the SNSA provision on an ongoing basis. Further improvements to the assessment system, ASN training and accessibility guidance are planned.

Practitioners who work with children and young people with complex additional support needs should use their knowledge and understanding of

the individual child and their needs, strengths and challenges to reach a decision, with parents, on whether or not the use of standardised assessment is appropriate.

What should you do now?

Supporting schools

Local authorities will decide how they support their schools with the SNSA. Local authority representatives with responsibility for assessment are working closely with headteachers and teachers to advise them on planning for the assessments. Advice and support on classroom management and on the use of assessment data is particularly important.

Using the data

All the data generated by the SNSA is owned by local authorities and provides them with the opportunity to analyse children's and young people's progression in a variety of ways.

Reporting and analysis of the data is key, and local authorities should give consideration to how they wish to produce reports that are useful and

Training

With SCHOLAR, all local authorities have negotiated training packages, incorporating both phases of training, appropriate to their own needs and at a time in the school session best suited to them.

To support the conduct of the norming and equating studies, schools have received guidance notes and a bespoke training package for teachers.

Technical readiness

The SNSA are entirely web-based and available on any technical platform. They do not require any bespoke IT upgrades or installations. However, local authorities need to ensure technical readiness in schools so that the assessments can be successfully implemented.

informative. Local authority level reports have been made live in the SNSA system and additional data has been issued separately. This provides local authorities with a wide range of SNSA data to use for analysis.

Curriculum for Excellence levels

Each local authority has its own assessment and moderation policy which should include strategies for considering SNSA data and how the results can be used diagnostically to inform learning and teaching. Consideration should also be made on how the SNSA support teachers' professional judgement on children's achievement of Curriculum for Excellence levels.

The SNSA Service Desk can provide help and support on all aspects of the SNSA including advice on technical readiness.

You can email the Service Desk at **info@snsa.org.uk (mailto:info@snsa.org.uk)**.Phone contact details for the Service Desk are available when you log into the SNSA system itself.