

# Case Studies

## National Monitoring of Education

One important function of assessment is to monitor national changes in the education of young people so that the various stakeholders, including educators and the public and their representatives, can take any necessary actions to improve the quality of education. The following case studies present two examples, the National Assessment of Educational Progress (NAEP) in the United States and the National Educational Monitoring Project (NEMP) in New Zealand. Following the descriptions of these two national assessments, Table 1 on page 36 compares the ways in which they meet (or do not meet) the assessment standards.

### *Case 1: The United States' National Assessment of Educational Progress*

The NAEP was developed as a test broad enough to cover what the designers considered appropriate educational domains including mathematics, reading, science, writing, the arts, civics, economics, geography, and U.S. history. The test, which was far too big for individual students to take, was then broken into smaller overlapping tests. These have been administered to a representative national sample of 9-, 13-, and 17-year-old students every four years since 1969. The four-year cycle was considered appropriate because shorter term systemic change was viewed as relatively unlikely.

The tests, which include multiple-choice and extended-answer questions, are administered by individuals hired and trained specifically for the purpose. The sampling system was designed to be nationally representative but is deliberately structured in such a way that comparisons cannot be made among states, school districts, or cities. Such comparisons were viewed as likely to increase the stakes involved and thereby encourage people to engage in activities such as “teaching to the test,” which would then affect the extent to which the results could provide a valid representation of general achievement.

The NAEP results are presented to the public as scaled scores (from 0–300 or 0–500, depending on the subject) and at five percentiles through National Report Cards. Gains and (particularly) losses in performance are attended to by the press and politicians. The numbers remain relatively abstract since only a small percentage of the items are released for scrutiny by the public. The item structure of this long-term trend assessment test has been consistent since 1971 so that direct comparisons can be made over time. Participation is mandatory

and sampling includes public and private schools, though in 2004 the private school sample was too small to be reported.

In 1990, politicians decided that enabling state-by-state comparisons would be a good idea, and energy was diverted to development of a second NAEP test. This second test, now called the “main NAEP,” is administered at grades 4, 8, and 12, only in public schools. It allows state-by-state comparisons and, on a trial basis, comparisons of large urban districts. It is administered every two years and changes about every ten years to reflect curriculum changes. Tests are administered in science, math, reading, and writing. They are all administered in English. Some students are excluded for various reasons. Although participation in the state-level test had been voluntary, the No Child Left Behind Act of 2001 required states receiving Title I money to participate in the reading and math tests. Test items include multiple-choice, extended-answer, and short-answer questions, and results are reported both in scaled score performance levels and in categories of achievement (basic, proficient, and advanced) determined by cut scores. These are reported to the public by the press, though it seems likely that most who receive the information have little idea of what is meant by either the scaled scores or the categories (i.e., what it means to be “proficient”).

### ***Case 2: New Zealand’s National Educational Monitoring Project***

NEMP uses a national sampling of students over four-year cycles to assess 15 different areas of the national curriculum: art, music, speaking, listening, viewing, health and physical education, science, reading, writing, math, information skills, graphs, tables and maps, social studies, and technology. Knowledge, skills, motivation, and attitudes are all assessed. The assessment includes items addressing material not in the school curriculum in order to monitor the effects of any changes in the national curriculum. Students are assessed in English at two pivotal transition periods, year 4 (age 8–9) and year 8 (age 12–13). In Māori Medium settings, assessment is only at year 8. There is a deliberate effort to accommodate a range of differences in language, culture, gender, ability, and disability in the design and administration of assessment tasks. There are virtually no exclusions.

Almost all items are performance based, requiring students to work on tasks for three to four hours spread over five days, with the support of a trained teacher–test administrator. Tasks are selected to be meaningful and enjoyable for the students to ensure optimal engagement and the best picture of their capabilities. The task formats include working one-on-one with the teacher–administrator, working cooperatively in a group of four, and working independently on a series of hands-on activities or pencil-and-paper tasks. Some of the activities are videotaped and scored with rubrics. All items are carefully piloted.

In the NEMP, literacy is viewed as a social activity as much as a cognitive activity. For example, one task has a group of four year 4 students acting as the library committee. They are given a set of books and must choose, individually and then collectively, which books the library should purchase. The videotaped event is scored for the collaborative process as well as for individual performance.

School participation is voluntary; if a school is selected on multiple occasions or is unable to participate in a given testing, it is replaced with the most comparable school available. Replacement is rare because of a history of positive experiences. The test is administered by a group of teachers who are seconded from the schools, trained, and then returned to their teaching after the six-week test-administration period. Teachers are involved in the development of tasks, trialing of items, administration of tasks, and analysis of responses, and they report that the experience provides excellent professional development, which they share with their schools upon their return.

Results are reported to the public and to educators in terms of national performance and the performance of subgroups by demographics (e.g., race, gender, school size and characteristics). Results are reported in different formats to accommodate a wide audience, but typically they are reported in concrete terms of types of item citing specific examples. About two thirds of the items are released in order to maintain transparency and, in addition, so that teachers might use these items to see how their students compare with the national sample.

**Table 1. Analysis of National Monitoring Cases 1 and 2 in Relation to the IRA–NCTE Assessment Standards**

<b>Assessment standard</b>	<b>Case 1: NAEP</b>	<b>Case 2: NEMP</b>
1. The interests of the student are paramount in assessment.	Relatively little attention is paid to student interests in generating items. Although one aspect of the test invites a broad curriculum, increasing pressures associated with the test have curriculum-distorting potential.	Items are selected for student engagement. Assessments are closely tied to professional development in order to improve teaching. The assessment addresses a broad curriculum without high stakes that would distort the curriculum.
2. The teacher is the most important agent of assessment.	Little attention is paid to the teacher's role.	Teacher professional development is deliberately linked to training for test administration and scoring tasks.
3. The primary purpose of assessment is to improve teaching and learning.	There is no deliberate link between assessment and the improvement of teaching and learning, though, in recent years, there has been increasing pressure for higher generic scores.	Teacher professional development is specifically linked to training for administration and scoring tasks. Teachers are able to use excellent items as part of their instruction.
4. Assessment must reflect and allow for critical inquiry into curriculum and instruction.	The curriculum is addressed broadly in the lower stakes long-term test but less broadly in the higher stakes "main NAEP," increasing the likelihood of curriculum distortion in the latter test.	The curriculum is addressed broadly; communication of results is extensive and in plain language with concrete examples, and the performance of subgroups is analyzed. Because items go beyond the current curriculum, the effects of changes in curriculum can be analyzed.
5. Assessment must recognize and reflect the intellectually and socially complex nature of reading and writing and the important roles of school, home, and society in literacy development.	Items represent literacy as an individual cognitive activity with a modest degree of complexity relatively unconnected to home and society. The test recognizes the value to schools and society of a broad description of the consequences of education.	Items reflect the full complexity of literacy in a wide range of contexts and applications, both individual and social. Tasks are drawn from in-school and outside-school practices and deliberately attend to cultural and linguistic matters.

(continued)

**Table 1. Analysis of National Monitoring Cases 1 and 2 in Relation to the IRA–NCTE Assessment Standards (continued)**

Assessment standard	Case 1: NAEP	Case 2: NEMP
6. Assessment must be fair and equitable.	Items are selected and piloted to ensure fairness. Only one language is represented. Test performances are analyzed to reveal educational inequities. Private (mostly religious) schools are not clearly represented in the sampling system.	Items are selected and piloted to ensure fairness. Both primary cultural languages are represented. Item performances are analyzed to reveal educational inequities. If a school selected through the sampling system declines to participate, another school with similar characteristics is selected to ensure representation.
7. The consequences of an assessment procedure are the first and most important consideration in establishing the validity of the assessment.	The initial intention of an untainted indicator of national educational efforts is no longer realized because of changes in the sampling system (both items and students) that invite distorting pressures.	The consequence of the procedure is primarily professional development for teachers and a wide awareness of goals and progress in schooling. Because high stakes are not attached to the testing there is little incentive to distort the curriculum. These were central considerations in the design of the assessment.
8. The assessment process should involve multiple perspectives and sources of data.	The preparation of the test specifically includes a wide range of cultural and stakeholder representatives.	The preparation of the test specifically includes a wide range of cultural and stakeholder representatives.
9. Assessment must be based in the local school learning community, including active and essential participation of families and community members.	Local involvement is encouraged mostly through distribution of test results. Few items are released, which limits the meaningfulness of results to members of the public.	Local involvement is encouraged mostly through distribution of results and education of the public by providing extensive examples of test items and examples of the range of responses. This process regularly reminds the public of the breadth of the curriculum.

*(continued)*

**Table 1. Analysis of National Monitoring Cases 1 and 2 in Relation to the IRA–NCTE Assessment Standards (continued)**

Assessment standard	Case 1: NAEP	Case 2: NEMP
10. All stakeholders in the educational community—students, families, teachers, administrators, policymakers, and the public—must have an equal voice in the development, interpretation, and reporting of assessment information.	Representatives of various stakeholder groups are engaged in the development and trialing of the assessment items.	Representatives of various stakeholder groups are engaged in the development and trialing of the assessment items, an ongoing process since a large percentage of items is released to the public. Representatives are also convened to discuss and interpret the results as part of releasing information to the press. The assessment program is politically independent, limiting the relative power of some otherwise more powerful groups.
11. Families must be involved as active, essential participants in the assessment process.	Families have access to assessment information about changes in the effects of schooling.	Families have access to concrete, interpretable assessment information about changes in the effects of schooling.

## School and Classroom Assessments: Response to Intervention in the United States

Beginning in 1975 in the United States, federal money was set aside for the education of children considered “handicapped.” Children considered handicapped because of their failure to learn to read or write were classified as learning disabled because of a discrepancy between expected achievement (on the basis of a measure of intelligence) and actual achievement on an academic test.

Several problems arose with this process. First, the number of children classified as learning disabled expanded enormously. Second, a disproportionate number of minority students were so classified. Third, it took an extended period before the discrepancy was considered sufficient for these children to be classified and receive the benefit of the financial resources set aside for them. In the reauthorization of the federal Individuals With Disabilities Education Act (IDEA), an alternative was introduced to address these problems. Fifteen percent

of the money allocated for special education could be used for intervention programs intended to prevent the need to classify children as learning disabled. The premise was that, before limited achievement could be assumed to be caused by a learning disability, instructional interventions should be attempted in order to rule out the possibility of inadequate instruction.

There are few actual requirements of the law. It requires that children's learning be monitored over time to determine whether instruction is effective ("data-based documentation of repeated assessments of achievement at reasonable intervals, reflecting formal assessment of student progress during instruction"). It requires that the instructional intervention is "scientific, research-based"—the definition of which is very broad. Finally, it requires that, in order to classify a child as learning disabled, there must be procedures and a committee (including the child's parents), a relevant classroom teacher, and "at least one person qualified to conduct individual diagnostic examinations of children."

Researchers and school districts have approached this in different ways. One family of Response to Intervention (RTI) approaches focuses on the use of intervention to identify students with learning disabilities. The other family of approaches focuses centrally on preventing students from needing to be classified as learning disabled. Examples of each are represented in these cases following, and a comparative analysis of the two is presented in Table 2 on page 42.

### ***Case 3: An Identification Focus***

One approach to implementing RTI involves screening children for potential difficulties using the Dynamic Indicators of Basic Early Literacy (DIBELS) to select those who are at risk of failure in reading. These children are given additional instructional attention. To ensure that no children who might need assistance are missed, children's reading progress is monitored from the middle of first grade on by measuring once each week how many words each child can read accurately from grade-level passages in one minute. The passages are standardized and norm-referenced, and reliability is emphasized. In kindergarten and the first half of second grade, progress is monitored by a measure of how quickly children can break a word into separate sounds and give a name and a sound for a letter. Trained aides, special education teachers, and the school psychologists complete most of the assessments in order to limit the testing time required of the classroom teacher. Students are given a comprehensive standardized reading test at the end of each year. Students who do not improve their reading speed and accuracy sufficiently, or at an adequate rate, after eight weeks are given a small-group instructional program taught by a trained teacher aide. Students who still do not increase their speed and accuracy receive an intensified intervention with increased time in a smaller group, taught by the literacy specialist.

These interventions are referred to as tiers, classroom instruction being tier 1, and successive interventions as tiers 2 and 3.

The instructional intervention is based on a program shown in experimental studies to be effective at increasing children's ability to read words with greater speed and accuracy according to the federal What Works Clearinghouse website. The program is a standardized intervention package with a set sequence of materials and a scripted instructional format. Fifteen minutes are spent on phonics, word recognition, and spelling regular and irregular words; five minutes on building speed with letter names, letter sounds, and word family patterns; and ten minutes reading short passages (3 to 4 words to over 40 words) based on sounds and words previously taught. During the ten minutes, comprehension questions integrating literal and inferential thinking are asked and strategies for locating answers are taught. Teachers are monitored by the school psychologist to ensure that they are implementing the program with fidelity—that is, as scripted.

Before initiating a new tier of intervention, a committee—directed by the school psychologist and including a classroom teacher, a parent (or surrogate), the principal, and a special education teacher—meets to decide whether the next phase is appropriate, given the assessments. Parents are kept informed using the graphs and norms of reading speed and accuracy. Those students who do not benefit from these interventions are referred by the committee to special education for individual instruction (tier 4) and classified as learning disabled. Failure to benefit from a validated form of instruction is seen as evidence of a learning disability.

#### ***Case 4: A Prevention Focus***

This approach to RTI also involves layers of instruction, screening, and monitoring. On entering kindergarten, children are screened for their knowledge of the alphabet, and those with limited knowledge are given extra instructional support from the start on the assumption that limited alphabet knowledge reflects a limited literate history. Progress is monitored with an agreed upon portfolio of indicators including dated pieces of writing; an alphabet record for recording cumulative knowledge of letters, sounds, and related words noticed during classroom learning and one-on-one conferences; records of children's reading processes (strategies and accuracy); book difficulty level data; and anecdotal records. Some of these data are replaced in first and second grade with rubrics for judging writing, including writing stemming from reading. Comprehension is assessed through book discussions (small group, large group, and individual).

These portfolios are examined in monthly collaborative grade-level meetings led by a literacy coach who has 20% of her time designated for such



administrative work. The coach is part of the school's commitment to improving instructional quality to reduce the need for additional interventions. At the end of each quarter, children's learning is evaluated at grade-level meetings led by the literacy coach and the principal in terms of end-of-grade expectations. These meetings include instructional planning.

The core classroom program has differentiated small-group instruction with the classroom teacher providing additional support for the lowest group. The school has a highly trained literacy coach who works with teachers 60% of her or his time to improve tier 1 instruction. Tier 2 is a small-group intervention with group size, amount of time, and teacher expertise determined by the students' needs, but with the framework consistent with tiers 1 and 3. Each is focused on interactions that support meaning making and independence. Tier 3 is a 1:1 intervention with Reading Recovery in first grade or a 1:2 group or reading/writing conferences in upper grades. (According to research and the federal What Works Clearinghouse website, Reading Recovery is a program shown in experimental studies to be effective at increasing children's ability to comprehend and to read and spell more accurately and to reduce the number of children becoming learning disabled.) The small-group interventions are carried out by Reading Recovery teachers and by special education teachers trained in the approach, and the literacy coach spends 20% of her or his time teaching these interventions. Tier 4 includes, as part of the referral process, close examination of the teaching interactions in tier 3 by an expert coach with collaborative attempts to change instructional interactions of students who are not adequately accelerating their ability to handle more difficult texts.

At the beginning of first grade, those children in the bottom half of the class are assessed using the Observation Survey of Early Literacy Achievement, a standardized procedure that offers instructionally useful information regarding literacy concepts, knowledge, and processes. This assessment is used to allocate students to tier 2 or 3. The intervention teachers keep daily records of writing, word work, and reading processes, and classroom teachers continue to accumulate portfolios of children's writing and running records of their reading. There is a comprehensive assessment at the end of each grade.

Before initiating a new tier of intervention, a committee, directed by the principal and the literacy coach and including the classroom teacher and a parent (or surrogate), meets to examine progress and next steps. Parents are kept informed of progress using half-year reports for all students and monthly descriptive feedback by intervention teachers using, for example, writing and text-level examples.

**Table 2. Analysis of School and Classroom Cases 3 and 4 in Relation to the IRA–NCTE Assessment Standards**

<b>Assessment standard</b>	<b>Case 3: Identification focus</b>	<b>Case 4: Prevention focus</b>
1. The interests of the student are paramount in assessment.	Instructional adaptations serve accurate diagnosis of disability and assumes that the student's interests are best served by identifying genuine and permanent handicaps so that subsequent accommodations can be made.	Instructional adaptations prevent initial difficulties from becoming disabilities and assumes that the student's interests are best served by attributing lack of progress to instructional inadequacies, prompting constant efforts at instructional improvement.
2. The teacher is the most important agent of assessment.	Teacher role is minimized in assessments by having others gather assessment data. Teacher role in intervention-as-assessment is also restricted by enforcing program fidelity, minimizing teacher adaptation for a particular child.	The teacher gathers ongoing formative data and individually and collaboratively negotiates instructional strategies based on those data. Teacher expertise is central in noticing, collecting, and responding to data in instruction/intervention. The emphasis on ongoing coaching recognizes this.
3. The primary purpose of assessment is to improve teaching and learning.	The focus of assessment is on reliably determining which students are not benefiting from instruction rather than on providing instructionally useful information. Data collected on teaching are not to improve instructional interactions but to ensure instruction is not influenced by individuals.	Data are collected by the teacher to ensure they inform instruction. Regular stock-taking meetings are to counter individual biases and problem-solve instruction for students not accelerating adequately. Data are gathered specifically at tier 4 on instructional interactions to improve teaching.
4. Assessment must reflect and allow for critical inquiry into curriculum and instruction.	Data on teaching only allow for standardizing instruction and pointing to students for whom instruction is not working. Data do not inform the nature of instructional improvement. Because the focus of assessment is narrow (speed and accuracy of word reading), the differential effects of the larger literacy curriculum cannot be examined.	Data are collected on both teaching and learning that allow inquiry into curriculum and instruction. Assessments address a wide array of literacy (word knowledge, writing, comprehension) as well as teaching interaction patterns, enabling critical inquiry into the curriculum and its effects.

*(continued)*

**Table 2. Analysis of School and Classroom Cases 3 and 4 in Relation to the IRA–NCTE Assessment Standards (continued)**

Assessment standard	Case 3: Identification focus	Case 4: Prevention focus
5. Assessment must recognize and reflect the intellectually and socially complex nature of reading and writing and the important roles of school, home, and society in literacy development.	This case does not recognize literacy as social or complex and involves parents in the process of classifying a student as learning disabled.	This case recognizes literacy learning as social and somewhat complex.
6. Assessment must be fair and equitable.	Fairness is approached as ensuring due process, equal treatment, and reliable data and for providing accommodations for those with handicaps.	Fairness is viewed as requiring optimal instruction for all, which might be different for each.
7. The consequences of an assessment procedure are the first and most important consideration in establishing the validity of the assessment.	Reliability is considered the foundation of validity. Validity is tied to a narrow view of literacy. A valid assessment is considered to be one that accurately identifies students who are, in fact, learning disabled and does not identify those who are not.	An assessment process is considered valid if it leads to optimal instruction and the prevention of learning disability.
8. The assessment process should involve multiple perspectives and sources of data.	Multiple perspectives may be represented at the committee meeting. However, since data are narrow, there is limited likelihood that different perspectives will be invoked.	Multiple perspectives can be represented at quarterly grade-level meetings and at committee meetings. A broad range of data are available to invite and address different perspectives.
9. Assessment must be based in the local school learning community, including active and essential participation of families and community members.	Assessment is based in the local school learning community with limited participation of families.	Assessment is based in the local school learning community with limited participation of families.

*(continued)*

**Table 2. Analysis of School and Classroom Cases 3 and 4 in Relation to the IRA–NCTE Assessment Standards (continued)**

Assessment standard	Case 3: Identification focus	Case 4: Prevention focus
<p>10. All stakeholders in the educational community—students, families, teachers, administrators, policymakers, and the public—must have an equal voice in the development, interpretation, and reporting of assessment information.</p>	<p>This standard is not sustained within this part of the school assessment system.</p>	<p>This standard is not sustained within this part of the school assessment system.</p>
<p>11. Families must be involved as active, essential participants in the assessment process.</p>	<p>Families are primarily involved at critical junctures.</p>	<p>Families are primarily involved at critical junctures.</p>