# USING THE RESULTS OF A NATIONAL ASSESSMENT OF EDUCATIONAL ACHIEVEMENT

**Thomas Kellaghan**
**Vincent Greaney**
**T. Scott Murray**

**Correlates of Achievement**

A national assessment usually collects information on demographic and other background factors to allow comparisons to be made between the achievements of subgroups in the population. This information, when related in statistical analyses to student achievement, can answer questions that are central to the role of a national assessment, such as the following:

– Is the system underserving any particular group?

– Are gaps between groups in performance large enough to warrant remedial action?

– What factors are associated with low achievement?

In this way, by relating outcomes to inputs that are provided and processes being used, a national assessment shows "what is." However, it can also show "what might be" by demonstrating that some sectors in the system are achieving more desirable outcomes (high achievement) and by attempting to identify the factors associated with relative success.

If sample sizes in a national assessment are sufficiently large, evidence can be provided on achievement by gender; region (for example, province); location (urban or rural); membership of ethnic or language groups; and type of institution attended (public or private). Prompted in part by the current emphasis on gender equity in the Education for All and Fast Track Initiative programs, national assessments usually examine achievement differences between boys and girls. Figure 2.2 summarizes Sri Lanka's national assessment results for the percentages of male and female students achieving "mastery" in their first language, mathematics, and English.

A national assessment in Kuwait also identified gender differences in reading achievements (figure 2.3). In this case, gender differences were associated with

---

***Текстът е продължение на статия от кн.1/2014 г. на сп. „Стратегии на образователната и научната политика“***

***The text started in issue 1/2014 of “Strategies for Policy in Science and Educational Journal”***

length of time watching television. Not surprisingly, perhaps, long periods spent watching television were associated with lower reading achievement scores.

Table 2.8 provides information on the mathematics achievements of racial groups in South Africa based on data collected for the Trends in International Mathematics and Science Study (TIMSS). Table 2.9 provides information on regional differences in a national assessment in Nepal.

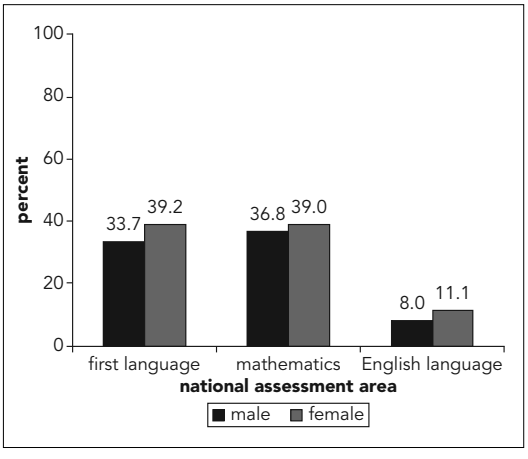


**Figure 2.2** Percentages of Male and Female Students Achieving Mastery in National Assessment, Grade 4, by Curriculum Area: Sri Lanka

*Source: Sri Lanka National Education Research and Evaluation Centre 2004: figure 4.44. Reproduced with permission.*

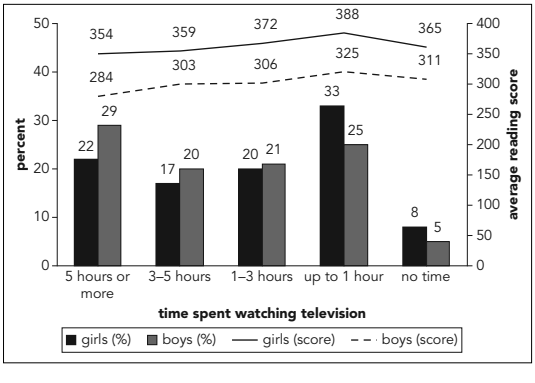


**Figure 2.3** Percentages of Students Watching Television for Varying Amounts of Time, by Gender and Average Reading Score: Kuwait

*Source: Kuwait Ministry of Education 2008. Reproduced with permission.*

Ethiopia's national assessment also provides information on the distribution of achievement by region (figure 2.4). Findings of the assessment suggest at least two policy options. They indicate, on the one hand, that support might be directed to the regions with the greatest number of low-scoring students (Oromiya, Somali, Dire Dawa) on a composite measure (based on scores on the reading, mathematics, environmental science, and English tests). On the other hand, the decision to provide support might be based on the lowest mean performance scores (Tigray, Benishangul-Gumuz).

**Table 2.8**
Mean Scores in Mathematics, Grade 8, by Racial Group, South Africa

| Racial group | Number of students | Mean score | Standard error | Minimum | Maximum |
|---|---|---|---|---|---|
| African | 5,412 | 254 | 1.2 | 5 | 647 |
| Asian | 76 | 269 | 13.8 | 7 | 589 |
| Colored | 1,172 | 339 | 2.9 | 34 | 608 |
| Indian | 199 | 341 | 8.6 | 12 | 612 |
| White | 831 | 373 | 4.9 | 18 | 699 |
| Total or overall mean[a)] | 8,147 | 275 | 6.89 | | |
| International mean | | 487 | 0.7 | | |

*Source: Howie 2002. Reproduced with permission.*

a) Based on South African national data set for the Third International Mathematics and Science Study Repeat.

**Table 2.9**
Mean Scores in National Assessment of Nepali Language, Grade 5, by Region: Nepal

| Region | Number | Mean scores | Standard deviation |
|---|---|---|---|
| Eastern | 802 | 51.32 | 16.7 |
| Central | 932 | 50.91 | 19.5 |
| Western | 1,018 | 52.89 | 13.2 |
| Midwestern | 465 | 50.78 | 12.7 |
| Far western | 293 | 49.71 | 13.2 |

*Source: Nepal Educational and Developmental Service Centre 1999. Reproduced with permission.*

Many (but not all) national assessments collect information in questionnaires about aspects of students' educational experiences and their home and community circumstances to provide clues about the school and extra-school factors that infl uence students' scholastic progress. In the Vietnam national assessment, for example, a private corner for study in a student's home was associated with higher achievement, even when more global assessments of home background were taken into account. High- and low-achieving schools and students could also be differentiated on the basis of regularity of meals and number of days absent from school (World Bank 2004).

Interpreting the findings of analyses in which student achievement is related to other variables requires some caution. A conclusion that factors related to student achievement may be considered to influence or "cause" achievement may not be warranted for a number of reasons, including the following:

– Causal interpretations of relationships identified in cross-sectional data can usually be sustained only if supported by other evidence.

– The number of schools or students in some categories may be too small to allow reliable inferences to be made.

– Methods of statistical analysis may be inappropriate. Analyses that explore relationships between two variables (for example, between class size and achievement) and that fail to take account of complex interactions in data (for example, between teacher qualifications and location of school) can lead to erroneous interpretations.

– Complex analyses are required to account for the influence on education outcomes of interacting factors operating at the student, school, and classroom levels. Because of this situation, multilevel, multivariate statistical techniques are required. These techniques isolate the effects of a variable that are net of the effects of other variables by systematically removing or adjusting for the effect of clusters of variables to show that significant differences exist or do not exist among students and schools after adjustment. For example, a conclusion that private schools are superior to public schools based on a finding that students attending private schools have higher levels of achievement than students attending public schools may not be supported when students' scores are adjusted to take account of their socioeconomic background.

Appreciation of the complexities involved in identifying causes of achievement will be strengthened by realizing that, more often than not, the background variables for which data are collected in a national assessment may only be proxies for factors that affect student learning at a deeper level and in a more diffuse way. In that case, other research data need to be considered; even sophisticated statistical analyses may not be adequate. For example, although analysis might reveal a positive correlation between student learning and the number of books in a student's home (see figure 2.5), one would not be justified – even when other variables are taken

into account – in concluding that number of books is causally related to student achievement. Although access to books may be important, student learning is likely affected not directly by the availability of books but by characteristics of an environment that cherishes books, such as one in which parents place a high value on scholastic achievement, provide academic guidance and support for children, stimulate children to explore and discuss ideas and events, and set high standards and expectations for school achievement (see Kellaghan and others 1993).

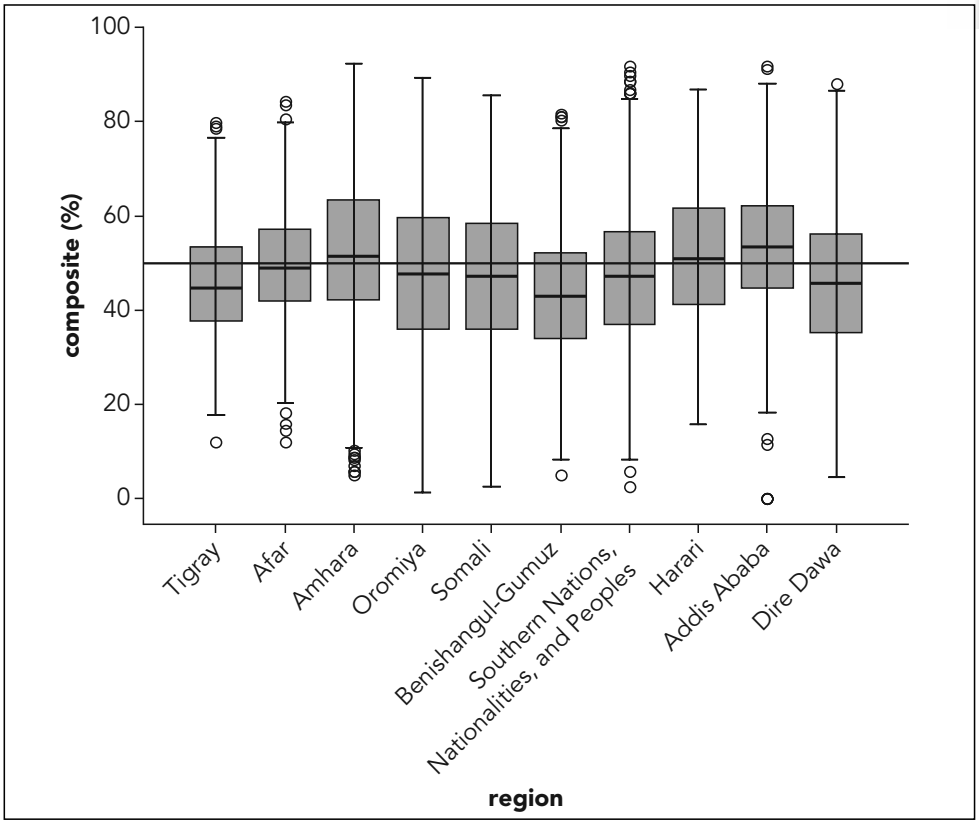


**Figure 2.4**
Regional Differences in Achievement, Grade 4: Ethiopia

*Source: Gebrekidan 2006. Reproduced with permission of the Ethiopia Quality Assurance and Examinations Agency.*

*Note: The mean is represented by the heavy black line; the box extends from −1 standard deviation (SD) to +1 SD, and the end lines from −1.96 SD to +1.96 SD. Markings outside the end lines represent outliers or extreme values.*

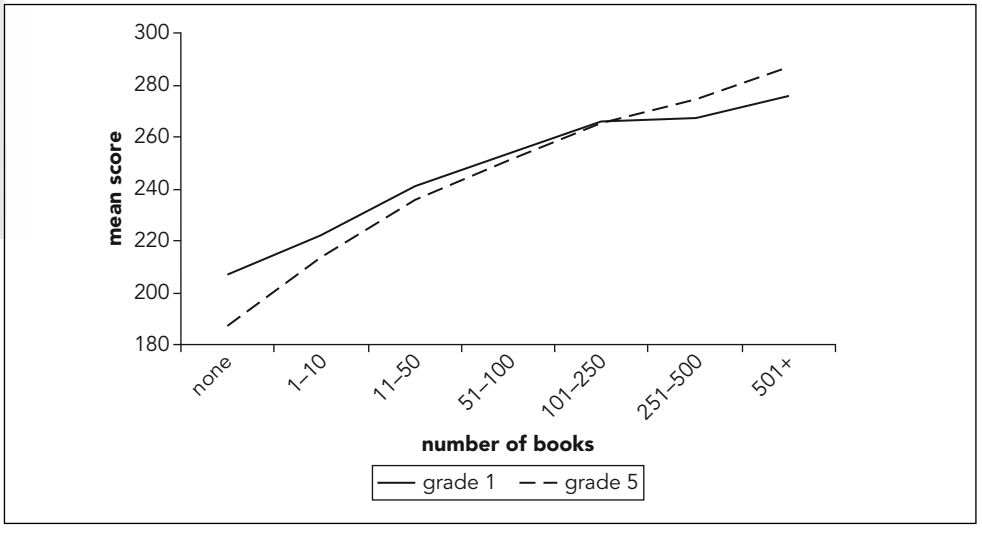


**Figure 2.5**
Mean Reading Test Scores of Students Plotted against Number of Books in the Home, Grades 1 and 5: Ireland
*Source: Eivers and others 2005a: figure 4.2.*

**Change In Achievement Over Time**

National assessments provide evidence of change in student achievement over time if assessment instruments are properly linked. When this information is available, findings may be presented as in the U.S. national assessment (NAEP) for the years 1971 to 2004 (figure 2.6). The plot of average reading-scale scores indicates that for students at age nine the average reading score was higher in 2004 than in any previous year. For students at age 13, the average score in 2004 was higher than the average score in 1971 and 1975, but no different from the average score in other years. Mean score in 2004 for students at age 17 showed a decrease from 1992. Volume 1 of this series (Greaney and Kellaghan 2008: 134) reported on changes in literacy scores over time in a number of African countries. If assessment instruments permit (that is, if they provide adequate representation of curriculum subdomains), national assessments can provide evidence of change not just in gross measures of student achievement, but also in subdomains of a curriculum area. For example, the tests used in grade 5 national assessments in Ireland allowed student performance to be estimated for a number of content strands and skills in mathematics. The results for assessments for five content strands, administered in 1999 and 2004, are presented in figure 2.7. The differences between years are

statistically significant only for "shape and space" and "data and chance." In both cases, an improvement was registered between 1999 and 2004.

Data from the same national assessment are reported for five mathematics skill areas in figure 2.8. In this case, only the increase for "reasoning" was significant between 1999 and 2004.

In many countries, the need to monitor progress toward achieving the Millennium Development Goals of universal primary education by 2015 reinforces interest in assessing achievement over time. Efforts to improve the quality of education, however, may have to proceed in the face of problems created by expanding enrollments and decreased budgets. In Malawi, for example, performance deteriorated sharply following a campaign to provide schooling for everyone without providing the necessary resources to deal with the abolition of school fees and increasing numbers (Altinok 2008).

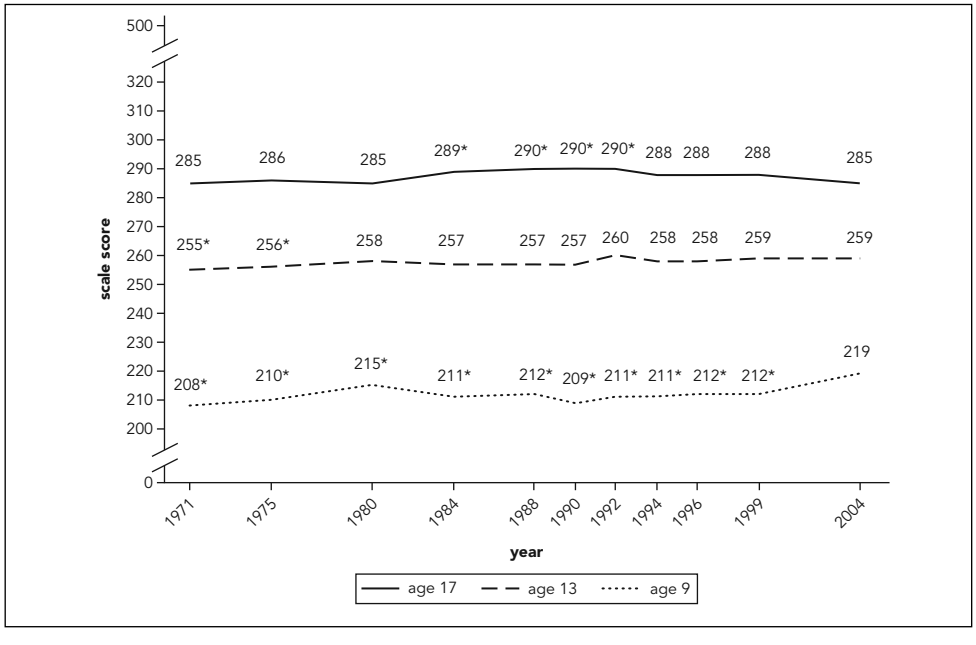


**Figure 2.6**
Trends in Average Reading Scale Scores for Students Ages 9, 13, and 17, NAEP, 1971–2004: United States

*Source: U.S. National Center for Education Statistics 2005.*

*Note: *indicates that a score differs signifi cantly from the score in 2004.*

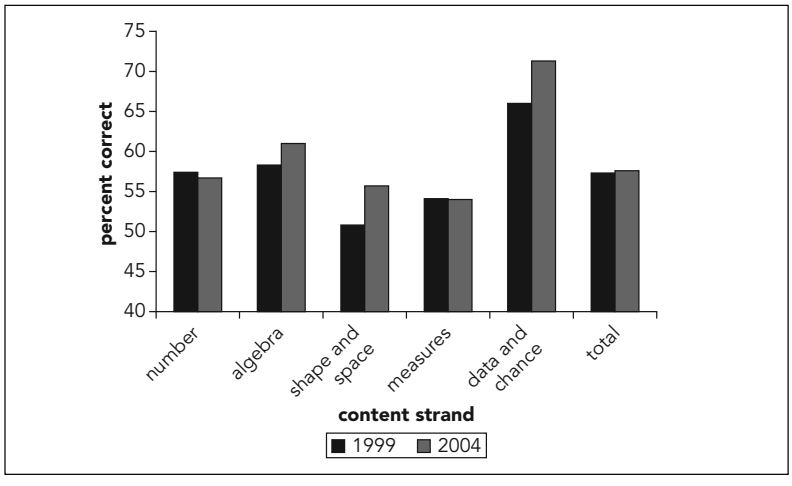


**Figure 2.7**
Mean Percent Correct Scores for Mathematic
*Source: Surgenor and others 2006: figure 3.1.*

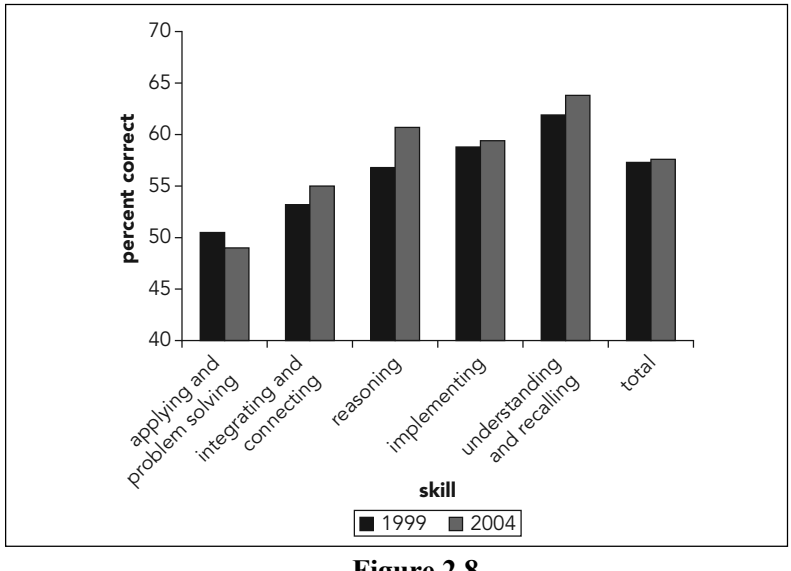


**Figure 2.8**
Mean Percent Correct Scores for Mathematics Skills in National Assessment, Grade 5, 1999 and 2004: Ireland
*Source: Surgenor and others 2006: figure 3.2.*

**Conclusion**

A main national assessment report should describe the study in sufficient detail to act as a primary source of information for the assessment. At the same time, it should not overload readers with inordinate amounts of technical information. Some reports do; others err in the opposite direction by providing inadequate information on technical aspects of the study, content of achievement tests, methods used, or error and bias in estimates.

Another error is to provide data only in tabular form. Because few potential users have the skills required to draw inferences from statistical tables, illustrating key findings in charts or graphs is a better method. Furthermore, a report that relies heavily on tabular presentation may fail to pay due attention to research findings that help explain relationships revealed in the assessment, such as why a variable (such as home background) is important and what is known about its influence on student achievement. This information, however, is critical if users are to appreciate the import of findings and are to be in a position to judge what action or actions may be required to address identified deficiencies.

Views and practices differ regarding the inclusion of recommendations in a national assessment report. Some national (and international) reports contain recommendations; others do not. National assessment teams should clarify at the outset whether a ministry of education expects a report to contain recommendations. Furthermore, if recommendations are expected, national assessment teams should ascertain whether they should be included only if they arise directly from the national assessment study or whether broader considerations, such as relevant research findings, may be taken into account.