2001

A Policy Makers Guide to Systemwide Assessment Programs

Margaret Forster
ACER, Forster@acer.edu.au

Follow this and additional works at: http://research.acer.edu.au/policy_makers_guides
Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
http://research.acer.edu.au/policy_makers_guides/2

This Book is brought to you by the ACER Archives at ACEReSearch. It has been accepted for inclusion in Policy Maker's Guides by an authorized administrator of ACEReSearch. For more information, please contact repository@acer.edu.au.
This publication is the result of research that forms part of a program supported by a grant to the Australian Council for Educational Research by State, Territory and Commonwealth governments. The support provided by these governments is gratefully acknowledged.

The views expressed in this publication are those of the author and not necessarily those of the State, Territory and Commonwealth governments.
## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>What are the purposes of systemwide assessment programs?</td>
<td>3</td>
</tr>
<tr>
<td>Why the interest in systemwide assessment programs?</td>
<td>4</td>
</tr>
<tr>
<td>What are some examples of systemwide assessment programs?</td>
<td>6</td>
</tr>
<tr>
<td>How are data collected?</td>
<td>7</td>
</tr>
<tr>
<td>How are data reported?</td>
<td>8</td>
</tr>
<tr>
<td>How are trends monitored?</td>
<td>20</td>
</tr>
<tr>
<td>How are data used to improve learning?</td>
<td>27</td>
</tr>
<tr>
<td>What concerns have been raised?</td>
<td>37</td>
</tr>
<tr>
<td>A case study</td>
<td>44</td>
</tr>
<tr>
<td>A checklist of considerations</td>
<td>46</td>
</tr>
<tr>
<td>Useful websites</td>
<td>48</td>
</tr>
<tr>
<td>End Notes</td>
<td>49</td>
</tr>
</tbody>
</table>
This guide provides policy makers with research-based information about systemwide assessment programs.

Good decision making at all levels of an education system is facilitated by easily accessible, relevant, and reliable information.

Many indicators provide useful input to educational decision making; but the most important indicators are those which address the central concern of education: the promotion of student learning.

Education systems monitor student learning—with the fundamental intention of promoting learning—by collecting, analysing and reporting student achievement data. Given that state, national and international achievement studies are both time consuming and expensive, it seems prudent to reflect on this effort:

What is the purpose of these programs?
How are data reported and used?
How can we ensure that data will provide evidence for informed decision making?

INTRODUCTION

The assessment programs considered in this handbook focus on the collection and analysis of systemwide information.

Two kinds of information usually are collected:

• data on student achievement in particular subject areas at particular grade levels; and

• limited background information (characteristics of students).

Achievement data usually are collected through standardised tests administered to either representative samples or entire cohorts of students. Background information is collected by means of questionnaires completed by students or teachers.
WHAT ARE THE PURPOSES OF SYSTEMWIDE ASSESSMENT PROGRAMS?

Systemwide assessment programs provide systematic and regular measures of student learning. They are designed to investigate and monitor the ‘health’ of an education system and to improve student learning by providing information to stakeholders at different levels of the system. They provide:

- policy makers with information to monitor standards over time, to monitor the impact of particular programs, and to make decisions about resource allocation; and
- schools (principals, councils) and teachers with information about whole school, class and individual pupil performance that they can use to make decisions about resource allocation and to support learning in the classroom.

Full-cohort programs provide:

- parents with information about their child’s progress to assist them to make decisions about the best ways to support their child; and
- students with information about their progress to assist them to take an active role in monitoring their own learning.

Programs that are not full cohort can provide this information for a limited number of students.

The purpose of a systemwide monitoring program

The purpose of British Columbia’s Foundation Skills Assessment (FSA) program is stated explicitly. The program is intended to:

- provide information to districts about the performance of their students in relation to provincial expectations and standards in order to assist districts to plan for improvement;
- provide information to the public about the performance of students provincially in relation to expectations and trends over time;
- measure the achievement of students in reading comprehension, first-draft writing, and selected components of numeracy;
- determine if there are any trends in student performance at the district and provincial levels; and
- determine if there are groups of students who under perform with respect to provincial standards.
WHY THE INTEREST IN SYSTEMWIDE ASSESSMENT PROGRAMS?

The management of an education system is a complex and expensive operation. If decisions are to be informed, then dependable information on educational outputs is required. Systemwide programs provide this information for system level monitoring and resource allocation.

Of increasing interest, however, is the role large-scale assessment programs can play as agents of reform and accountability—to provide both direction and motivation to schools, teachers, parents and students. In the State of South Australia, for example, assessment is seen as the missing link in earlier curriculum planning and programming which was not informed, as a matter of course, by student achievement information. In British Columbia, State assessments are seen as part of the ongoing process of educational reform as Figure 1 below illustrates.

In some countries, student achievement data collected through systemwide assessment programs are used as a measure of schools’ contributions to student learning. In others, public comparisons of schools’ achievements in the form of ‘league tables’ are made.

[In the United States] the testing enterprise in K-12 education has mushroomed in the last quarter-century; Americans want numbers when they look at students, schools, state education systems, and how America’s students compare to those of other countries. Among political leaders, testing is turning into a means of reform, rather than just a way of finding out whether reforms have been effective.

In Figure 1 below illustrates a continuous process of assessment and educational reform.
United States government publications emphasise the need for system driven accountability measures to improve student learning:

No school improvement can succeed without real accountability for results. Turning around low-performing schools requires that state and district leaders take active steps to set high expectations for schools and students, establish the means to measure performance against those expectations, and create policies to identify and provide assistance to those schools and students that fail to meet high standards for performance.5

Commentary on this reform and accountability agenda focuses on its understandable political appeal:

Compared with reforms such as targeting instructional time, professional development for teachers, and reducing class sizes, state assessment programs are relatively inexpensive. The assessments also can be mandated (unlike changes in classroom practice), can be rapidly implemented, and have a public visibility.6

Significant commentary also focuses on the unintended negative consequences of assessment driven reform (see pages 38-44).

The recent role of achievement data in assessment driven reform is illustrated in the lower feedback loop in Figure 2. Once assessments of student learning (data collection) have been reported and evaluated (figure centre), information is then disaggregated to provide evaluations at school level. These evaluations are publicised. Rewards and sanctions are applied to schools to encourage improvements in student learning.

The upper feedback loop in Figure 2 illustrates the traditional approach to system use of student achievement data. Once assessments of student learning (data collection) have been reported and evaluated, decisions are then made at system level about the best ways to improve student learning through the allocation of resources to disadvantaged schools, to programs, and to teacher professional development.

**Figure 2 Using achievement data for improvement and accountability**
The table below lists systemwide assessment programs referred to in this handbook. Education systems collect a range of achievement data at key year levels of schooling. Most monitoring systems collect data on literacy and numeracy achievement, some, on a range of other learning outcomes.

### SOME EXAMPLES OF SYSTEMWIDE ASSESSMENT PROGRAMS

<table>
<thead>
<tr>
<th>Program</th>
<th>Year level/s assessed</th>
<th>Learning outcomes assessed</th>
<th>Website information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Australia</td>
<td>3, 5</td>
<td>Literacy, numeracy</td>
<td><a href="http://www.bos.vic.edu.au">www.bos.vic.edu.au</a></td>
</tr>
<tr>
<td>Victoria</td>
<td>3, 5, 7</td>
<td>Literacy, numeracy, science, studies of society and environment in alternate years</td>
<td></td>
</tr>
<tr>
<td>Queensland</td>
<td>3, 5</td>
<td>Literacy, numeracy</td>
<td><a href="http://www.qscc.quld.edu.au">www.qscc.quld.edu.au</a></td>
</tr>
<tr>
<td>Tasmania</td>
<td>3, 5</td>
<td>Literacy, numeracy</td>
<td><a href="http://www.tased.edu.au">www.tased.edu.au</a></td>
</tr>
<tr>
<td><strong>Canada</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brit. Columbia</td>
<td>4, 7, 10</td>
<td>Reading, numeracy, first draft writing</td>
<td><a href="http://www.bced.gov.bc.ca/assessments">www.bced.gov.bc.ca/assessments</a></td>
</tr>
<tr>
<td><strong>United States</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td>4, 8</td>
<td>Reading, Lang Arts</td>
<td><a href="http://www.ced.state.co.us/">www.ced.state.co.us/</a></td>
</tr>
<tr>
<td>Kentucky</td>
<td>4, 5, 7, 8</td>
<td>maths</td>
<td><a href="http://www.kde.state.ky.us/">www.kde.state.ky.us/</a></td>
</tr>
<tr>
<td>Nth Carolina</td>
<td>4, 8</td>
<td></td>
<td><a href="http://www.dpi.state.nc.us">www.dpi.state.nc.us</a></td>
</tr>
<tr>
<td>Texas</td>
<td>4, 8</td>
<td></td>
<td><a href="http://www.tea.state.tx.us">www.tea.state.tx.us</a></td>
</tr>
<tr>
<td>Tennessee</td>
<td>4, 8</td>
<td></td>
<td><a href="http://www.state.tn.us/education/">www.state.tn.us/education/</a></td>
</tr>
</tbody>
</table>
HOW ARE DATA COLLECTED?

Full-cohort testing
Some education systems collect student achievement data through full-cohort assessment programs. The Year 3 and Year 5 literacy and numeracy assessments in Australian States and Territories, in many states of the United States, and in parts of Canada are examples, although in some instances 'full-cohort' means the cohort of government schools only.

Typically, data in these programs are collected through the use of machine-scored paper and pen tests, although for some learning outcomes, students’ extended responses or on-the-spot performances are assessed by classroom teachers or by centrally trained assessors (for example, writing outcomes, and physical education outcomes). Data are aggregated to provide summaries of group performances, either at the level of a system or at the level of a school.

An advantage of these programs is that they can provide reliable information to all parents on individual student progress in a few crucial areas of school learning.

Sample surveys
Some systems collect achievement data through sample surveys which are designed to provide summary information at the system level only. The US National Assessment of Educational Progress (NAEP) is an example of a survey of this kind conducted at a national level. But some systems conduct similar surveys at a State or district level; for example, the Western Australian Monitoring Standards in Education Program.

Programs of this kind are based on the performances of carefully drawn representative samples of students. Samples may be drawn to ensure adequate representation of particular categories of students so that the average performances of students in those categories can be compared and monitored.

Although sample surveys cannot provide all parents with information on the progress of individual students, or local school communities with information on school results, sample surveys have a number of important advantages over full-cohort testing.

Sample surveys are capable of providing evidence about a rich and varied set of learning goals. Full-cohort testing programs inevitably address only those outcomes that can be assessed for many thousands of students at a time. This constraint limits the range of learning outcomes that the program is able to address.

Because sample surveys usually do not report on individual students, it is not necessary for all students to attempt the same set of assessment tasks. Different students can attempt different but overlapping sets of tasks (known as a 'multiple-matrix' design) to allow system reporting on a wide range of valued curriculum goals.

Sample surveys are also less expensive overall (though more expensive per student) than full-cohort testing, and tend to be less intrusive into classroom time (though they may require extensive commitment from a very small number of teachers and students).
HOW ARE DATA REPORTED?

There are multiple audiences for reports of student achievement, including education systems, school communities, parents, students, and the general public. These audiences usually are interested in different levels of detail about educational achievements.

Data are reported in summary or disaggregated form, and against a range of reference points; for example, curriculum standards, proficiency scales, and expectations. In each instance, findings are communicated using a variety of techniques including graphs, numerical presentations (tables), written descriptions, and rankings (‘league tables’).

The table below summarises some ways in which data are reported.

The following examples 1-10 illustrate reports of systemwide results. Examples 11-13 illustrate reports of school results, and example 14 illustrates an individual student report.

**SOME WAYS IN WHICH ACHIEVEMENT DATA ARE REPORTED**

**Systemwide results**
- Averages and distributions
- Against national norms
- Against standards framework (including described proficiency scales)
- Against performance expectations (including standards, goals, benchmarks)
- Against international benchmarks
- For subgroups of students (including gender, cultural background, language background)
- Against background variables
- In curriculum areas
- Item-by-item

**School results**
- Averages
- For subgroups of students
- For subgroups of schools (including district, geographic region)
- In curriculum areas
- Item-by-item
- Against systemwide cohort achievement
- Against school level performance goals (including value-added)

**Student results**
- Individual student results
Example 1  Systemwide—averages and distributions

Most education systems provide summary reports of the mean achievements of students. In this example, the summary achievements of Year 3, Year 7 and Year 10 Western Australian students in writing are shown graphically and (cover the page) in tabular form.

The performance scale (0-800), against which achievement is reported, is marked out in levels. A description of achievement at each of these levels is provided to the left of the display.

Level 8
Students strive to write with assurance, precision and vitality. They explore complex themes and issues in a variety of styles that compel readers’ interest and attention.

Level 7
Students explore ideas about texts and issues in a precise and organised way. They express themselves precisely when writing for complex purposes and they try to match text type, structure, tone and vocabulary to their subject and purpose.

Level 6
Students write in a variety of sustained ways to explore complex issues and ideas. They select information to influence readers. They make their meaning clear for readers by using correct punctuation, spelling and grammar and by manipulating words and the structure of the text.

Level 5
Students use a variety of text types to write at length and with some sense of complexity. They write sustained, logically organised texts that substantiate or elaborate ideas. They show a sense of the requirements of the reader and experiment with manipulating prose for effect.

Level 4
Students have a sound basic knowledge of how to use English. They use familiar ideas and information in their writing, showing control over the way some basic text types are written. They present ideas logically with limited elaboration. They try to adjust their writing to meet readers’ needs.

Level 3
Students write longer texts using ideas and information about familiar topics for particular purposes known audiences. They use many of the linguistic structures and features of a small range of text types and make attempts at spelling new words according to spelling patterns and conventions.

Level 2
Students produce brief written texts understood by others and which include related ideas and information about familiar topics. Students have a beginning knowledge of the conventions of written texts.

Level 1
Students show an emerging awareness of the nature, purposes and conventions of written language. They experiment with using written symbols for conveying ideas and messages.
Example 1 cont.....

<table>
<thead>
<tr>
<th></th>
<th>Number of students</th>
<th>Mean (in level)</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1992</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 3</td>
<td>1426</td>
<td>285</td>
<td>2</td>
</tr>
<tr>
<td>Year 7</td>
<td>1497</td>
<td>477</td>
<td>4</td>
</tr>
<tr>
<td>Year 10</td>
<td>1143</td>
<td>564</td>
<td>5</td>
</tr>
<tr>
<td><strong>1995</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 3</td>
<td>1682</td>
<td>262</td>
<td>2</td>
</tr>
<tr>
<td>Year 7</td>
<td>1610</td>
<td>475</td>
<td>4</td>
</tr>
<tr>
<td>Year 10</td>
<td>1563</td>
<td>551</td>
<td>5</td>
</tr>
</tbody>
</table>

Example 2  Systemwide— against national norms

Many State and district education systems report systemwide summary achievement against national norms. For example, many States of the US use National Assessment of Educational Progress (NAEP) data and TIMSS results as reference points for student achievement on State tests. The table below shows results on the Connecticut Mastery Test 1999 compared with a national sample of students.

State level normative information

<table>
<thead>
<tr>
<th>Grade</th>
<th>Mean national percentile ranking *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mathematics</td>
</tr>
<tr>
<td>Grade 4</td>
<td>68</td>
</tr>
<tr>
<td>Grade 6</td>
<td>66</td>
</tr>
<tr>
<td>Grade 8</td>
<td>68</td>
</tr>
</tbody>
</table>

*Normative information is provided to indicate how well the average student in Connecticut performs compared to a United States national sample. For example, it is estimated that fourth grade students who achieved the state average score on the CMT mathematics test would have scored better than 68% of students nationally.
### Example 3  Systemwide—against standards frameworks

Increasingly, systems are reporting summary achievement against levels of a standards framework. This example shows the summary reporting of Grade 4 students’ reading achievement against National Assessment of Educational Progress (NAEP) proficiency scales.⁹

<table>
<thead>
<tr>
<th>North Carolina</th>
<th>NAEP State Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994 Reading Grade 4</td>
<td>1996 Math Grade 4</td>
</tr>
<tr>
<td>Proficient level and above</td>
<td>Proficient level and above</td>
</tr>
<tr>
<td>30%</td>
<td>21%</td>
</tr>
<tr>
<td>Basic level and above</td>
<td>Basic level and above</td>
</tr>
<tr>
<td>59%</td>
<td>64%</td>
</tr>
</tbody>
</table>
Example 4  Systemwide—against standards frameworks (proficiency scales)

As part of its Monitoring Standards in Education sample assessment program, Western Australia assesses and reports the speaking achievements of Year 3, 7 and 10 students. This example shows the achievements of Year 3 students in small group discussion against described speaking proficiency levels. The achievements are shown graphically. Four levels of the 8-level proficiency scale are described below.

**Speaking proficiency levels**

**Level 8**
Students use group work to explore complex concepts. They negotiate agreements in groups where there are disagreements or conflicting personalities, managing discussions sensitively and intelligently and concluding with positive summaries of achievement.

**Level 7**
Students make significant contributions to independent work groups and are aware of and able to modify their own behaviour where necessary. They attempt to arrive at consensus. Students discuss the topic at a sophisticated level.

**Level 6**
Students explore ideas in discussions by comparing their ideas with those of their peers and building on others’ ideas to advance the discussion. They generate a comprehensive and detailed response to the topic.

**Level 5**
Students work well in formal groups where they take on roles, responsibilities and tasks. They consider challenging issues, give considered reasons for opinions and ideas, and constructively discuss the presentation of those ideas.
Example 5  Systemwide—against performance expectations

Where systems have set proficiency standards (the level at which students are expected to perform), summary and disaggregated data often are reported as the percentage of students achieving the proficiency standard.

This example shows the Grade Reading/Language Arts achievements of Grade 3 students in Oregon. The first row shows the achievement of the Year 3 cohort (based on eighty-eight per cent participation rate). The second and third rows show the achievements of students from low socio-economic backgrounds. Federal government Title 1 funds provide additional educational support for children in need. Schools with more than fifty per cent poverty are eligible for a ‘school wide’ program. Targeted assistance programs provide funds directly to the neediest students. The entries at the bottom of the table show the achievements of students from a non-English speaking background.

<table>
<thead>
<tr>
<th></th>
<th>% Standard not met</th>
<th>% Meet standard</th>
<th>% Exceed standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>All students</td>
<td>23.0</td>
<td>42.0</td>
<td>35.0</td>
</tr>
<tr>
<td>Title 1 school wide</td>
<td>32.0</td>
<td>44.5</td>
<td>23.5</td>
</tr>
<tr>
<td>Title 1 targeted</td>
<td>50.2</td>
<td>42.4</td>
<td>7.4</td>
</tr>
<tr>
<td>Percent of school in poverty</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-34%</td>
<td>17.0</td>
<td>40.0</td>
<td>43.0</td>
</tr>
<tr>
<td>75-100%</td>
<td>46.0</td>
<td>40.0</td>
<td>14.0</td>
</tr>
<tr>
<td>Limited English Proficient students</td>
<td>72.3</td>
<td>25.5</td>
<td>2.2</td>
</tr>
<tr>
<td>Migrant students</td>
<td>53.8</td>
<td>41.6</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Example 6  Systemwide—against international benchmarks

Although the age/grade groups of students involved in national and international assessments are often different from State and district target grades, in some systems there is enough overlap to be of interest. For example, many States of the US use Third International Mathematics and Science Study (TIMSS) results as a reference point for student achievement on State tests.

British Columbia, as part of its State assessment report, details student achievement on the national School Achievement Indicators Program (SAIP) and in 1998 included 1995 TIMSS results.
Example 7  Systemwide— for subgroups of students

Most systems report the achievements of subgroups of the population, including breakdowns by gender, cultural and language background, geographic region, and type of community.

The first graph shows the achievements of students from a language background other than English (LBOTE), Aboriginal and Torres Strait Islander students (ATSI), boys, girls and all students in Year 5 Mathematics, Western Australia.12

The second graph shows the achievements of Grade 4 students in different regions of Newfoundland and Labrador on the complete battery of Canadian Basic Skills tests.1

The third graph shows the reading and viewing achievements of Year 5 students in Queensland, Australia—proportionally more girls are in the top scoring groups.14
Example 8  Systemwide—against background variables

Some systems collect and report information about students’ attitudes and behaviours.

British Columbia, for example, reports data on students’ reading, writing and mathematics attitudes and behaviours, including calculator and computer use, at Grades 4, 8 and 10. Changes in attitudes and behaviours as students move from Grade 4 through to Grade 10 are of particular interest. In Grade 4, 68% of students reported they read in their spare time almost every day. The percentage of students reading in their spare time every day dropped to 51% at Grade 7, and 35% at Grade 10.

Example 9  Systemwide— in curriculum areas

Some systems report achievement by curriculum area within test. This example shows the achievements of students in South Australia on different aspects of numeracy. Included are the achievements of boys and girls. The reported means are based on scaled scores (not raw scores) allowing for the comparison of achievements on sub-sets of test items. For example, in 1997, students performed less well in ‘space’ than in ‘number’ or ‘measurement’.

<table>
<thead>
<tr>
<th>Year 5 1997 Numeracy aspects: Number, Space, Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>All students</td>
</tr>
<tr>
<td>58.3</td>
</tr>
<tr>
<td>58.5</td>
</tr>
<tr>
<td>56.4</td>
</tr>
<tr>
<td>Boys</td>
</tr>
<tr>
<td>58.4</td>
</tr>
<tr>
<td>59.0</td>
</tr>
<tr>
<td>56.4</td>
</tr>
<tr>
<td>Girls</td>
</tr>
<tr>
<td>58.3</td>
</tr>
<tr>
<td>58.0</td>
</tr>
<tr>
<td>56.5</td>
</tr>
</tbody>
</table>
Example 10 Systemwide—item-by-item

It is also possible to report the performances of groups of students by item. For example, Queensland reports the achievements of boys, girls, Aboriginal and Torres Strait Islander students, students from language backgrounds other than English, and different groups of language speakers on each test item. The table below shows the performance of students from an English speaking background (ESB), students from a non-English speaking background (NESB) with English as their first language (non-ESL), and students from a non-English speaking background (NESB) with English as a second language (ESL), on a number of 1998 Year 5 numeracy items. On these items, students from an NESB + ESL background performed better than their NESB and ESB peers. Apart from item 40, all of these items are from the 'Number' strand and demand calculation skills of varying levels of difficulty.

<table>
<thead>
<tr>
<th>Item number and item description</th>
<th>% answering correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ESB</td>
</tr>
<tr>
<td>01 Add amounts of money</td>
<td></td>
</tr>
<tr>
<td>(calculator available)</td>
<td>82.0</td>
</tr>
<tr>
<td>07 Calculate using the correct</td>
<td></td>
</tr>
<tr>
<td>order of operations (calculator</td>
<td>50.5</td>
</tr>
<tr>
<td>available)</td>
<td></td>
</tr>
<tr>
<td>08 Calculate using the correct</td>
<td></td>
</tr>
<tr>
<td>order of operations (calculator</td>
<td>40.9</td>
</tr>
<tr>
<td>available)</td>
<td></td>
</tr>
<tr>
<td>10 Calculate the remainder of a</td>
<td></td>
</tr>
<tr>
<td>division example (calculator</td>
<td>38.2</td>
</tr>
<tr>
<td>available)</td>
<td></td>
</tr>
<tr>
<td>20 Multiply a 2 digit number by a</td>
<td></td>
</tr>
<tr>
<td>single digit number involving</td>
<td>61.5</td>
</tr>
<tr>
<td>regrouping</td>
<td></td>
</tr>
<tr>
<td>21 Subtract 3 digit numbers</td>
<td></td>
</tr>
<tr>
<td>with regrouping</td>
<td>68.1</td>
</tr>
<tr>
<td>38 Interpret a table and order</td>
<td></td>
</tr>
<tr>
<td>lengths involving decimals</td>
<td>20.7</td>
</tr>
<tr>
<td>40 Interpret visual information</td>
<td></td>
</tr>
<tr>
<td>and solve problem involving</td>
<td>5.8</td>
</tr>
<tr>
<td>distances around a 3D shape</td>
<td></td>
</tr>
</tbody>
</table>
Example 11 School results—averages

In many systems the performances of individual schools are reported publicly. For example, in Pennsylvania the mean performance of schools on the Pennsylvania System of School Assessment (PSSA) sorted by county and district is available on the Department of Education website. Average scaled scores for schools in Brownsville district are shown below.

### Average scaled scores for Year 5

<table>
<thead>
<tr>
<th>County</th>
<th>District</th>
<th>School</th>
<th>Math</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fayette</td>
<td>Brownsville</td>
<td>Cardale El Sch</td>
<td>1170</td>
<td>1190</td>
</tr>
<tr>
<td>Fayette</td>
<td>Brownsville</td>
<td>Central El Sch</td>
<td>1220</td>
<td>1300</td>
</tr>
<tr>
<td>Fayette</td>
<td>Brownsville</td>
<td>Colonial El Sch</td>
<td>1220</td>
<td>1260</td>
</tr>
<tr>
<td>Fayette</td>
<td>Brownsville</td>
<td>Coox-Donahey El Sch</td>
<td>1170</td>
<td>1210</td>
</tr>
<tr>
<td>Fayette</td>
<td>Brownsville</td>
<td>Hiller El Sch</td>
<td>1200</td>
<td>1250</td>
</tr>
</tbody>
</table>

Example 12 School results— for subgroups of students

Many systems report information directly to schools. This school report is provided to schools in the Australian Capital Territory. It shows whole school and subgroup achievement in listening at Year 3 level.

### SCHOOL REPORT

<table>
<thead>
<tr>
<th>Year 3 Assessment in Listening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole School (No 63)</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

Average

- Whole School: 13.1
- Boys: 13.8
- Girls: 13.5
Example 13 School results— against standards frameworks

The following report is provided to schools in South Australia. The report shows the school’s results against the State proficiency scales (Bands).²⁰

Basic Skills Testing Program 1995 Year 5 Aspects of literacy
School:  ##
No of students:  96

Percentage of students in skill bands

<table>
<thead>
<tr>
<th>Band</th>
<th></th>
<th>Reading</th>
<th>Language</th>
<th>Literacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>State</td>
<td>45</td>
<td>37</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>School</td>
<td>52</td>
<td>36</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>State</td>
<td>22</td>
<td>32</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>School</td>
<td>23</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>State</td>
<td>18</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>School</td>
<td>16</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>State</td>
<td>15</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>School</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>
Example 14 Student results—individual student achievement

Some full cohort assessment programs provide individual student reports. This Queensland Year 6 report shows an individual student’s numeracy results on Level 2, Level 3 and Level 4 items and against systemwide performances.21

<table>
<thead>
<tr>
<th>Number</th>
<th>Measurement</th>
<th>Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Represents common fractions on a number line.</td>
<td>Convert measurements using common metric prefixes.</td>
<td>Visualise position and describe it on a map using distance, direction and co-ordinates.</td>
</tr>
<tr>
<td>Use place value to compare and order numbers.</td>
<td>Use conventional units of mass.</td>
<td>Make use of distance relating to co-ordinate pairs.</td>
</tr>
<tr>
<td></td>
<td>Calculate areas of rectangles.</td>
<td>Visualise locations and understand directional language when reading maps.</td>
</tr>
</tbody>
</table>

**Number**

- Level 4
  - Completely solve division problems by interpreting remainders.
  - Divide a whole number by a 1-digit number.
  - Subtract a 3-digit number from another involving regrouping.
  - Recognise equivalent fractions.
  - Continue number patterns.
  - Place whole numbers in order.
  - Subtract one 3-digit number from another.
  - Partly solve division problems by interpreting remainders.
  - Multiply by a 1-digit number.
  - Interpret whole numbers written in words and use a calculator for adding whole numbers.
  - Represent word problems as number sentences.
  - Use a calculator for subtracting whole numbers.

- Level 3
  - Use a calculator to add lengths expressed as decimals.
  - Compare areas by counting units.
  - Use an object as a repeated unit of measurement.
  - Identify a right angle.
  - Calculate time intervals.
  - Choose shapes that can cover a region with no gaps or overlaps.
  - Compare and measure length to the nearest graduation.

- Level 2
  - Recognise that different units can be used to measure the same length.
  - Locate a date on a calendar.
  - Read time on a clock.
  - Read a thermometer scale to the nearest marked graduation.

- Level 1
  - Recognise features of a 3D object that are shown in a 2D diagram.
  - Recognise the same shape within arrangements and patterns.
  - Recognise 3D shapes from a description of their surface.
  - Select a flat shape that will fold to make a prism.

**Measurement**

- Represent common fractions on a number line.
- Use place value to compare and order numbers.
- Convert measurements using common metric prefixes.
- Use conventional units of mass.
- Calculate areas of rectangles.

**Space**

- Visualise position and describe it on a map using distance, direction and co-ordinates.
- Make use of distance relating to co-ordinate pairs.
- Visualise locations and understand directional language when reading maps.
- Visualise and follow paths using co-ordinates.
- Recognise features of a 3D object that are shown in a 2D diagram.
- Recognise the same shape within arrangements and patterns.
- Recognise 3D shapes from a description of their surface.
- Select a flat shape that will fold to make a prism.

- Compare properties of 3D shapes.
- Interpret placement of objects in drawings.
Systems monitor the ‘health’ of an education system by studying trend data. Some ways in which trends are monitored are summarised in the table below. Examples 15–23 provide detailed illustrations.

**SOME WAYS IN WHICH TRENDS ARE MONITORED**

**Statewide trends**
- Changes in averages and distributions
- Changes in percentage of students above or below national norms
- Changes in percentage of students at levels of a standards framework
- Changes in percentage of students above or below performance expectations
- Changes in the achievements of subgroups of students (including relative growth of subgroups)

**School trends**
- Changes in averages and distributions
- Changes in the achievements of subgroups of students
- Changes in school rankings ('league tables')
- Success in meeting performance goals (including value added)
Example 15  Systemwide trends—averages and distributions

The Australian Capital Territory monitors student achievement by comparing the median achievements and distributions over time and between year levels.22
Example 16 Systemwide trends—percentages in levels

One of the ways in which South Australia monitors student achievement over time is by tracking the percentage of students working within a ‘skill band’ on a described proficiency scale. The graph illustrates the percentage of Year 3 students achieving skill bands 1, 2, 3, 4 and 5 in numeracy in 1995, 1996, and 1997. In 1997, the drop in the mean in Year 3 was accompanied by an increase of 2% of students in skill bands 1 and 2, and a 5% decrease in skill bands 4 and 5. There was a 4% increase of students in skill band 3.

Example 17 Systemwide trends—reaching performance expectations

Connecticut monitors student achievement by comparing over time the percentages of students at or above State goals set for each grade. Goals are set for each grade on each of three tests (mathematics, reading comprehension, and written communication). For example, the State goal for reading at each grade level is set at 8 on a scale of 2–12. The table shows the percentage of students who performed at or above the State goal on all three tests.

<table>
<thead>
<tr>
<th>Year</th>
<th>Grade 4</th>
<th>Grade 6</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>19.3</td>
<td>23.1</td>
<td>21.5</td>
</tr>
<tr>
<td>1994</td>
<td>23.3</td>
<td>23.7</td>
<td>25.4</td>
</tr>
<tr>
<td>1995</td>
<td>27.1</td>
<td>24.3</td>
<td>28.4</td>
</tr>
<tr>
<td>1996</td>
<td>30.1</td>
<td>30.0</td>
<td>36.5</td>
</tr>
<tr>
<td>1997</td>
<td>32.8</td>
<td>30.2</td>
<td>36.4</td>
</tr>
<tr>
<td>1998</td>
<td>34.9</td>
<td>33.8</td>
<td>40.4</td>
</tr>
<tr>
<td>1999</td>
<td>34.5</td>
<td>38.1</td>
<td>41.5</td>
</tr>
<tr>
<td>Change</td>
<td>+15.2</td>
<td>+15.0</td>
<td>+20.0</td>
</tr>
</tbody>
</table>
Example 18  Systemwide trends—reaching performance expectations

Western Australia tracks the percentage of students at or above the fitness standards, for different ages, established in 1994 by the Australian Council for Health, Physical Education and Recreation (ACHPER). In 1998, the state reported a decline with age in the percentage of students who achieved the appropriate ACHPER minimum standard of cardiorespiratory endurance. This decline was greatest for girls, with 60% of Year 10 girls not achieving the minimum standards for their age.25

![Cardiorespiratory Endurance Chart](chart.png)

Example 19  Systemwide trends—reaching performance expectations

Connecticut also tracks the percentage of students, over the three test administrations, scoring above the goals set for each grade. The table shows the reading achievements of three different cohort groups from 4th to 6th to 8th grade.26

<table>
<thead>
<tr>
<th></th>
<th>Grade 4</th>
<th></th>
<th>Grade 6</th>
<th></th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>year</td>
<td>%</td>
<td>year</td>
<td>%</td>
<td>year</td>
</tr>
<tr>
<td>Cohort 1</td>
<td>1993</td>
<td>44.6</td>
<td>1994</td>
<td>59.4</td>
<td>1995</td>
</tr>
<tr>
<td>Cohort 2</td>
<td>1994</td>
<td>45.0</td>
<td>1995</td>
<td>60.0</td>
<td>1996</td>
</tr>
<tr>
<td>Cohort 3</td>
<td>1995</td>
<td>47.7</td>
<td>1996</td>
<td>60.3</td>
<td>1997</td>
</tr>
</tbody>
</table>
Example 20 Systemwide trends—subgroups of students

Western Australia tracks the mean achievements (on the Monitoring Standards in Education performance scale) of students in Years 3, 7 and 10 in writing. The table shows mean writing scores for subgroups of the population.27

Summary of subgroup performances in writing 1995

<table>
<thead>
<tr>
<th></th>
<th>Year 3</th>
<th>Year 7</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>262</td>
<td>475</td>
<td>551</td>
</tr>
<tr>
<td>Girls</td>
<td>285</td>
<td>502</td>
<td>583</td>
</tr>
<tr>
<td>Boys</td>
<td>238</td>
<td>450</td>
<td>517</td>
</tr>
<tr>
<td>ATSI*</td>
<td>140</td>
<td>372</td>
<td>448</td>
</tr>
<tr>
<td>NESB**</td>
<td>229</td>
<td>428</td>
<td>521</td>
</tr>
</tbody>
</table>

*Aboriginal and Torres Strait Islander students

** Students from a non-English speaking background

Example 21 Systemwide trends—subgroups of students

Queensland tracks the mean achievements of students in Years 3, 5 and 6 to investigate the gender gap. The table below shows that in literacy (reading and viewing, spelling, and writing) girls consistently outperform boys, with the gap being widest at Year 5 and reduced at Year 6.28 Boys appear to catch up most noticeably during Year 6 in reading and viewing.

Mean scores literacy

<table>
<thead>
<tr>
<th>Reading and viewing</th>
<th>Cohort</th>
<th>Boys</th>
<th>Girls</th>
<th>Gender gap in favour of girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998 Year 3</td>
<td>490.7</td>
<td>482.6</td>
<td>499.1</td>
<td>+16.5</td>
</tr>
<tr>
<td>1998 Year 5</td>
<td>592.6</td>
<td>582.6</td>
<td>602.8</td>
<td>+20.2</td>
</tr>
<tr>
<td>1997 Year 6</td>
<td>649.3</td>
<td>643.0</td>
<td>656.1</td>
<td>+13.1</td>
</tr>
<tr>
<td>1996 Year 6</td>
<td>629.5</td>
<td>621.5</td>
<td>638.5</td>
<td>+17.0</td>
</tr>
</tbody>
</table>
School trends

Some education systems disaggregate State achievement data to evaluate achievement at the school level with the aim of monitoring and comparing the contribution individual schools make to pupils’ progress.

At the crudest level, school averages are compared (and sometimes reported in the form of ‘league tables’). However, schools and their context differ from one another and these differences significantly influence progress and achievement. Raw results can be misleading indicators of the value added by a school if they are not adjusted for intake differences.30

Value added approaches

In response to this concern a number of ‘valued added’ approaches have been investigated. Value added is the calculation of the contribution that schools make to pupils’ progress.31 Taking a value added approach means finding an accurate way to analyse performance which takes account of factors that have been found empirically to be associated with performance but over which schools have little or no control. These factors include prior attainment, sex, ethnic grouping, date of birth, level of special education need, and social disadvantage. Typically, pupils’ prior attainment plus the overall level of social disadvantage in the school (as measured by free school meals) can account for as much as 80% of the apparent difference between schools.32

The term ‘value added’ also has been used more broadly (and sometimes confusingly) to describe a whole range of connected but distinct activities including

- making ‘like with like’ comparisons of schools’ (or departments’ or classes’) performances;
- representing pupils’ progress as well as their achievement;
- identifying which schools/departments/classes are currently performing above or below predictions; and
- identifying which individual pupils are likely to perform above or below predictions.33

Example 22 Systemwide trends—subgroups of students

By tracking the same cohort of students, South Australia monitors the relative growth of subgroups of the population. This graph illustrates numeracy growth from Year 3 to Year 5.29 The scale 0–9 represents growth on the calibrated tests. The largest growth occurs for the youngest students, but as this is a small group of students the results are relatively unreliable. Aboriginal students show the most growth with the oldest students showing least growth.
Example 23 School trends—value added

In Chicago, the value a school adds to students’ learning is calculated on the basis of a set of grade profiles.\textsuperscript{34} For each school, for each grade, the profile is based on two pieces of information: input status and learning gain. Input status captures the background knowledge and skills that students bring to their next grade of instruction and is based on students’ test scores on the Iowa Test of Basic Skills (ITBS) from the previous spring. Learning gain is the degree to which the end-of-year ITBS results have improved over the input status for the same group of students.

The profile is organised around data from a base year. Students who move into and out of a school during the academic year do not count in the productivity profile for that year. A statistical model is used to smooth trend lines so that variability in the data from year to year does not obscure any overall pattern. The model also is used to adjust trend estimates for other factors that might be changing over time besides school effectiveness (for example, the school’s ethnic composition, the percentage of low-income students, and retention rates).

To evaluate schools’ contributions to student learning, productivity profiles are classified into one of nine patterns using a dual indicator comparison scheme which considers both the learning gain trends and the output trends. For example, a school whose grade profile shows an increasing output trend with an input trend of the same rate is classified as ‘no change’ and contributing less to students’ learning than a school with an output trend which is increasing at a faster rate than the input trend (‘up’). In computing profiles for each grade, averages are calculated from across adjacent grades, providing a more stable estimate than single grade averages would provide. (Improving productivity in one grade tends to be followed by some declines in the next, and vice versa.)

Figure 3 shows the reading productivity trends for Fillmore Elementary School. The bottom trend line illustrates input trends; the top illustrates output trends. The distance between the two trend lines illustrates productivity.

Figure 3 Reading productivity for Fillmore elementary school

Note: Percentages associated with each grade productivity profile are the percentage improvement in learning gains over the base year period (1991).
Grade productivity profiles from all individual elementary schools in the system also are aggregated to show the overall productivity of Chicago schools. Figure 4 below illustrates overall ITBS mathematics productivity. The output trends are up for all grades and learning gain trends show improvements for the middle and upper grades. Grades three and four show little change in learning gains. The grade three data are particularly interesting. Although the output trend is positive, the Learning Grade Index is down by 4%. This indicates that the gains in achievement at the end of the grade are largely attributable to improvements prior to grade three. If output trend only had been considered, then one might have mistakenly concluded that third grades were improving system wide.

Figure 4  Mathematics productivity profile for Chicago public schools, 1987–1996

Note: LGI = Learning Gain Index, computed for 1992-1996

HOW ARE DATA USED TO IMPROVE LEARNING?

Student achievement data collected through systemwide programs usually are used for two closely related purposes: accountability and improvement.

For example, as part of their commitment to equal opportunity, education systems monitor the achievements of students from different geographic, gender and ethnic backgrounds to ensure that all students enjoy equal access to education. On the basis of achievement data, they may allocate additional resources to programs targeting a particular subgroup of students (system accountability to the public to provide resources equitably).

Systems that have set improvement goals check progress towards system targets. On the basis of achievement data, they may allocate additional resources to programs targeting low-achieving schools (system improvement purpose).

In some countries system managers also encourage schools to use system level achievement data for accountability and improvement. Schools are supported to use data to compare the achievements of their students with past performances, and with the performances of students in other schools, and to set school goals for improvement.
In some countries, schools are required to respond to centrally collected achievement data. For example, in twenty-three states in the US, schools are held directly accountable for education outcomes. These States have policies for intervening and mandating major changes in schools judged to be low performing on the basis of student achievement on State mandated tests. In some cases, States or districts provide technical assistance and additional resources to help redesign or restructure chronically low performing schools. In some jurisdictions, schools have been reconstituted. This often involves replacing school principals and removing teachers.35

The table below summarises some ways in which assessment data are used by system managers for improvement and accountability purposes at different levels of the education system. These strategies are elaborated (pages 29-37) with detailed examples. Example 24 illustrates one Canadian province’s systematic approach to using data at all levels of the education system to improve student learning.

<table>
<thead>
<tr>
<th>SOME SYSTEM USES OF ASSESSMENT DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement</td>
</tr>
<tr>
<td>System</td>
</tr>
<tr>
<td>• Allocating resources</td>
</tr>
<tr>
<td>• Motivating research</td>
</tr>
<tr>
<td>• Informing curriculum and</td>
</tr>
<tr>
<td>performance standards</td>
</tr>
<tr>
<td>School</td>
</tr>
<tr>
<td>• Providing professional development</td>
</tr>
<tr>
<td>• Setting goals</td>
</tr>
<tr>
<td>• Allocating resources</td>
</tr>
<tr>
<td>Classroom</td>
</tr>
<tr>
<td>• Providing curriculum feedback</td>
</tr>
<tr>
<td>• Motivating change</td>
</tr>
<tr>
<td>Student</td>
</tr>
<tr>
<td>• Informing learning</td>
</tr>
<tr>
<td>• Motivating learning</td>
</tr>
<tr>
<td>Parent</td>
</tr>
<tr>
<td>• Communicating progress</td>
</tr>
<tr>
<td>• Informing decision making</td>
</tr>
<tr>
<td>(school selection)</td>
</tr>
<tr>
<td>Pre-service Training</td>
</tr>
<tr>
<td>• Informing course focus</td>
</tr>
<tr>
<td>Educational Research</td>
</tr>
<tr>
<td>Community</td>
</tr>
<tr>
<td>• Informing about standards</td>
</tr>
</tbody>
</table>
Example 24 A systematic approach to using achievement data

The ways in which assessment data are to be used at all levels of the education system are made explicit in British Columbia. Information is used

by the Province and Districts to

• report on the results of student learning in selected areas of the curriculum;
• assist in policy, program and curriculum development;
• facilitate public and professional discussions on student learning;
• analyse results of particular populations of students to determine if they require additional support or focused attention;

by schools to

• facilitate discussions about student learning;
• assist in the development and review of school growth plans;

by students and parents as

• an additional external source of information about a student's performance in relation to provincial standards.

Example 25 System level—directing resources

Student achievement data from systemwide achievement studies can be used as a basis for allocating resources.

For example, in California the Immediate Intervention Underperforming Schools Program allocates additional resources to schools scoring in the bottom half of the Statewide distribution of the Standardized Testing and Reporting (STAR) Program. Schools may volunteer or may be randomly selected for planning grants to work with an external evaluator and a community team to identify barriers to school performance and to develop an action plan to improve student achievement.

In Queensland, State sector schools with students performing in the bottom 15% of the cohort on the literacy or numeracy Statewide tests are allocated additional funds to provide intervention programs designed specifically for those students.
Example 26 System level—motivating research

Student achievement data from systemwide achievement studies can be used to motivate research. For example, the literacy achievements of Queensland students relative to the achievement of students in other Australian States motivated two reviews—a study of the literacy practices in schools, and a study of the State testing program.

Example 27 System level—informing standards

Some systems use student achievement data to inform reviews of curriculum and performance standards. For example, the Victorian (State) Board of Studies commissioned a study to compare the English, science and mathematics expectations contained in the revised Curriculum and Standards framework with State and Territory data as well as with international achievement data. The intention was to confirm the level of expectation with reference to actual student performance.\(^3\)\(^8\)

Example 28 School level—providing professional development

Some systems provide direct assistance to schools to encourage them to pursue data driven improvements. For example, the Maryland State Department of Education has a web page to help schools to analyse their State data. Achievement data on a variety of key dimensions are presented in simple graphs for each school. The data are disaggregated by subject, gender, race, and grade. Schools can compare results to similar schools in the State. Worksheets also are provided to guide schools to investigate instructional practices, chart data, and identify further data they need to collect.\(^3\)\(^9\)

Similarly, in the State of Victoria, Australia, schools can use data to compare their students’ results with State and ‘like’ school benchmarks to learn about their effectiveness. They can compare their current performance levels with their own past performance, and the performance of similar schools, to plan for improved achievement and to set performance expectations for themselves.\(^4\)\(^0\)
Example 29 Classroom level—providing curriculum feedback

Data from full cohort State and district programs can be used to provide feedback to schools and teachers on student achievement in relation to aspects of the curriculum. On the basis of objective information, teachers are then able to adjust their teaching strategies. For example, the Australian Capital Territory provides schools and teachers with student achievement data on each test question. The report below has been annotated to assist teachers to see the kinds of ‘listening’ questions the students in this school found most difficult.41

---

### Annotated example — diagnostic School report

#### Year 5 Listening Item

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Skill Area</th>
<th>Item Description</th>
<th>School Report</th>
<th>NCS Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.2</td>
<td>Listens to sustained material with little distracting information</td>
<td>10%</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>16.3</td>
<td>Recognizes statements of attitude and opinion even when they are not directly stated</td>
<td>24%</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>16.4</td>
<td>Identifies the main ideas and supporting details from short spoken text</td>
<td>18%</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>22.2</td>
<td>Selects main idea and note the down in point form</td>
<td>22%</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>22.3</td>
<td>Uses and applies a system to record information quickly</td>
<td>16%</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>22.4</td>
<td>Listens to and retains the main idea of information from sustained spoken texts</td>
<td>14%</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>22.5</td>
<td>Listens to and retains a piece of information from sustained spoken texts</td>
<td>15%</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>22.6</td>
<td>Listens to and retains a piece of information from sustained spoken texts</td>
<td>17%</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>22.7</td>
<td>Listens to and retains a piece of information from sustained spoken texts</td>
<td>11%</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>22.8</td>
<td>Listens to and retains a piece of information from sustained spoken texts</td>
<td>9%</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>22.9</td>
<td>Listens to and retains a piece of information from sustained spoken texts</td>
<td>10%</td>
<td>29</td>
<td></td>
</tr>
</tbody>
</table>

### Performance on Listening Item

#### Annotated example — diagnostic School report

---

41 The items are scored from the most difficult to the least difficult, based on a performance of a national sample of students. Nis shown and more than one answer appears on each page on the test.42

---
British Columbia provides similar information to assist teachers to interpret district assessment results. A series of tables indicating the proportions of students answering each test question correctly are provided. Also provided is the proportion of students who selected particular incorrect alternatives and commentary (where possible) on what students are typically doing when answering questions incorrectly.

The Grade 4 results for ‘patterns and relations’ are listed below. The State numeracy assessment addresses number, patterns and relations, shape and space, and statistics and probability skills.

<table>
<thead>
<tr>
<th>Item</th>
<th>% correct</th>
<th>Description of item</th>
<th>Comments on incorrect responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>53</td>
<td>A word problem involving division and finding a remainder (dividing a number of objects into sets of a given size and finding how many are left over)</td>
<td>More than one quarter, 28%, subtracted rather than divided. 13 per cent divided correctly but found an incorrect remainder</td>
</tr>
<tr>
<td>21</td>
<td>55</td>
<td>A word problem involving multiplication, subtraction and division (finding the greatest number of an item which can be purchased with change from an earlier purchase)</td>
<td>Common errors were incorrect calculations, 19%, ignoring part of the information, 10%, and using only part of the information in the calculations, 10%</td>
</tr>
</tbody>
</table>
| 25   | 72        | Find the number of missing votes in a class of students using information shown in a tally | }
Encouraging the use of data at all levels of the system— a case study

British Columbia engaged an ‘interpretation panel’ of representatives of educational and community organisations to review and comment on the results of the 1999 State assessments. As well as commenting on strengths and areas requiring attention, the panel made recommendations regarding steps that could be taken to improve BC students’ reading comprehension, writing and numeracy skills.

Strategies were suggested at every level of the education system. Some of the recommendations are listed below:

**To teachers, principals, and superintendents**
- increase the amount of direct instruction in reading from Kindergarten to Grade 12
- emphasize that all teachers should teach reading strategies in all subject fields, not just in English Language Arts
- select reading materials that will engage boys
- encourage students to write daily
- increase emphasis on applications of mathematics and problem solving
- develop math intervention programs similar to reading intervention programs

**To the ministry**
- increase access to ESL programs
- provide updated writing reference sets (a classroom assessment resource) with samples of writing from provincial assessments
- provide additional support for implementation of mathematics Integrated Resource Packages

**To parents and guardians**
- encourage children to read regularly and read to your children regularly
- promote numeracy and problem solving in the home
- emphasize to children the importance of mathematics in our lives

**To teacher education programs**
- require that all education students take at least one course on the teaching of reading
- require that all education students take a course in the teaching of numeracy

**To educational researchers**
- increase research on strategies for teaching boys to read
- look into how different types of learners acquire reading, writing and numeracy skills
- research effective support strategies for numeracy
- conduct research on the relationship between student achievement and computer use/physical activities
Example 30 School level—accountability sanctions and rewards

Using State and national achievement data to hold schools accountable for improving student learning is of increasing interest in developed countries.

Twenty three States in the US have policies for intervening and mandating major changes in low performing schools, and 17 States grant this authority at district level. In some cases, this means that States or districts provide technical assistance and additional resources to help redesign or restructure chronically low performing schools. In some jurisdictions, schools have been constituted which often involves replacing school principals and removing teachers. For example, in Kentucky low performing schools are assigned ‘distinguished educators’ from other districts to assist in reform efforts. Schools that continue to drop far behind expectations are assigned state managers who evaluate all school personnel and make recommendations and changes to improve school performance.44

Subgroup achievement in Texas

The US State of Texas disaggregates student achievement data to measure both schools’ progress and the progress of students of different racial, ethnic and economic backgrounds.45 To make adequate yearly progress, schools must obtain an ‘acceptable’ rating from the State’s accountability system—a rating which requires at least forty per cent of all students and student groups to pass the Texas Assessment of Academic Skills, a dropout rate of no more than six per cent, and an attendance rate of at least ninety-four per cent. School districts can be disenfranchised and principals removed if sustained levels of performance are poor.

Achievement targets in Kentucky

The US State of Kentucky also has established a clear definition of adequate progress as part of the State accountability system.46 Student performance on assessments is classified: novice, apprentice, proficient, distinguished. Each classification is assigned a score (0, 40, 100, 140). The performance index for the school is defined as the average of these scores. A target score of 100 is set as a goal to be achieved within 20 years by all schools, and schools are expected to move a tenth of the way from their baseline performance toward the goal of 100 each biennium. For example, adequate progress for a school with a baseline index score of 30 would have a goal of 37 after two years (ie 30 plus 10% of 70—the difference between the baseline of 30 and the long term target of 100).

Schools that exceed the goals are eligible for financial awards and schools that fall behind are designated ‘in decline’. The lowest performing schools, ‘schools in crisis’, are those whose performance declines by more than five per cent of their baseline for two consecutive assessment cycles.
**School accreditation in Colorado**

New Colorado education accreditation indicators include academic achievement indicators. For example, schools are expected to increase the percentage of 4th grade students scoring at the proficient level or higher by 25% within three years. A district at 40% Proficient or Advanced would need to improve to 50% level within three years.47

**Awards and recognition in North Carolina**

The North Carolina Accountability Model for schools establishes growth/gain standards for each elementary, middle, and high school in the State. Schools that attain specified levels of growth/gain are eligible for incentive awards or other recognition.48 To be eligible for incentive awards, schools must not have excessive exemptions and must test at least 98% of their eligible students in K-8, and at least 95% of students enrolled in specific courses or grades in high school.

For example, ‘Schools of Excellence’ make expected growth/gain and have at least 90% of their students performing at or above grade level. They are recognised in a Statewide event, receive a dated banner to hang in the school, and a certificate. In addition, they receive whatever incentive award they earn as having made either ‘expected’ or ‘exemplary’ gains. Schools making exemplary growth receive a certificate and financial awards of $1500 per person for certified staff and $500 per person for teacher assistants.
Example 31 School level—accountability to parents and the public

Providing parents and the wider public with information about school achievement is of increasing interest in some countries. The most publicised strategy is to present results in the form of ‘league tables’. As well as providing information to parents and the public, it is assumed that leagues tables will generate competition and stimulate higher educational performance across the system.

Some States are experimenting with other forms of public accountability. For example, in 2000, the Australian Capital Territory Department of Education reviewed how information about students’ literacy and numeracy achievements was presented to parents and what information should be made publicly available. They presented five models for consideration including publishing information about average school results on Statewide tests, the distributions of students’ results on Statewide tests for each school, the performance of schools over time in relation to a base year, schools’ results against literacy and numeracy standards, and the progress of groups of students through school.

For example, using the last model, information that indicates the extent of improvement in student performance from Years 3 to 5 and 7 to 9 is published. This enables the comparison of the rate of improvement across schools as well as that of individual schools.

The intention is that this information will give parents, carers and the community an indication of progress over time, and whether mainstream and intervention school programs are actually making a difference to students’ learning. Student movement between schools needs to be taken into account to provide accurate data.

<table>
<thead>
<tr>
<th>Year 3</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>School System</td>
<td>School System</td>
</tr>
<tr>
<td>1997</td>
<td>1999</td>
</tr>
</tbody>
</table>
Over recent decades, a great deal has been learned about the ways in which large-scale assessment programs convey values and impact on practice, and about the unforeseen and unintended consequences of particular approaches to assessment. With the increasing emphasis in some countries on the use of student achievement data for accountability as well as improvement purposes, a new set of concerns has arisen.

There is a general concern about the emphasis placed on test scores:

In mandating tests, policy makers have created the illusion that test performance is synonymous with the quality of education.51

Technocratic models of school reform threaten to turn accountability into a narrow, mechanistic discussion based on numbers far removed from the gritty reality of classrooms.52

And, consequently, recommendations have been made to consider multiple indicators of performance:

Don't put all of the weight on a single test when making important decisions about students and schools (i.e., retention, promotion, probation, rewards). Instead, seek multiple indicators of performance. Include performance assessments and other indicators of success such as attendance, students taking Advanced Placement courses, etc.53

Specific concerns raised through systematic research about the reporting and evaluation of systemwide assessment data are summarised in the following table and discussed in detail below.

---

**Example 32 School level—holding students accountable**

Some States of America are using achievement results to hold students more accountable. In an effort to end 'social promotion', a number of States require districts and schools to use State standards and assessments to determine whether students can be promoted. For example, in Chicago, students who perform below minimum standards at key transition grades must participate in a seven-week summer bridge program and pass a test before moving on to the next grade. In 1997 about half of the 41,000 students who were required to attend the summer program passed the test. They showed an average one-and-a-half-year gain in their reading and mathematics scores.50

**WHAT CONCERNS HAVE BEEN RAISED?**

Over recent decades, a great deal has been learned about the ways in which large-scale assessment programs convey values and impact on practice, and about the unforeseen and unintended consequences of particular approaches to assessment. With the increasing emphasis in some countries on the use of student achievement data for accountability as well as improvement purposes, a new set of concerns has arisen.

There is a general concern about the emphasis placed on test scores:

In mandating tests, policy makers have created the illusion that test performance is synonymous with the quality of education.51

Technocratic models of school reform threaten to turn accountability into a narrow, mechanistic discussion based on numbers far removed from the gritty reality of classrooms.52

And, consequently, recommendations have been made to consider multiple indicators of performance:

Don't put all of the weight on a single test when making important decisions about students and schools (i.e., retention, promotion, probation, rewards). Instead, seek multiple indicators of performance. Include performance assessments and other indicators of success such as attendance, students taking Advanced Placement courses, etc.53

Specific concerns raised through systematic research about the reporting and evaluation of systemwide assessment data are summarised in the following table and discussed in detail below.
Over interpreting improvement in test scores

The problem: It is possible for test scores to go up without an increase in student learning in the domain the test addresses. This can happen if teachers teach to a non-secure test. What came to be known as the ‘Lake Wobegon Effect’ (below) is an example of inflated impressions of student achievement.

Example 33 The Lake Wobegon Effect—inflated impressions of student achievement

The mushrooming of standardised tests started in the US in the 1970s with ‘minimal competency’ testing. By 1987 John Cannell, a physician in West Virginia, noticed that many States and schools were claiming that their students were reported as being above average. An investigation revealed that students’ scores almost everywhere were above average. Cannell concluded that ‘standardized, nationally normed achievement tests give children, parents, school systems, legislatures and the press inflated and misleading reports on achievement levels’.

In his assessment of Cannell’s concerns, Robert Linn summarised his response:

There are many reasons for the Lake Wobegon effect...among the many are the use of old norms, the repeated use of the same test year after year, the exclusion of students from participation in accountability testing programs at a higher rate than they are excluded from norming studies, and the narrow focusing of instruction on the skills and question types used on the test.
A solution: Assessments can lead teaching without a negative impact if the assessments are standards-referenced—that is, if the tests are systematically constructed to address publicly available standards. Teachers then teach to the standard, not to specific test items.56

The problem: Rises (or falls) in test scores, which make politically attractive headlines, may be insignificant due to sampling or measurement error, or based on invalid comparisons—where there have been changes in the testing format, administration (eg testing time allowed), or exclusion policies.

For example, trend results from States and districts in the US that include a shift from an old to a new test show that where a new test replaces one that has been in use for several years there is a sudden drop in achievement. This drop is followed by a steady improvement (‘sawtooth’ effect).57

A solution: When monitoring trends over time, it is important to report measurement errors and to ensure that like comparisons are made. In general, for monitoring purposes, the group mean is a more reliable statistic than the percentage of students achieving a particular performance standard or working at a particular level of a standards framework.

Over interpreting systemwide trend data

The problem: When systemwide trends are being monitored over time, average score increases at the year levels tested are sometimes observed. These increases do not necessarily indicate that schools are performing better over time—for example, there may be no increase over time in cohort growth between the year levels tested.

Disaggregating summary data by school can also give a distorted picture of growth. A NAEP study (example 34) illustrates these problems in the context of national monitoring but the same issue arises in the context of other systemwide monitoring programs.

Example 34 NAEP—over interpreting summary trend data

A redesign in NAEP in the early 1980s allowed the tracking of a cohort of students, in addition to measuring the level of 4th, 8th and 12th grade students at a given time. In most cases, the students’ average NAEP scores were slightly higher at each grade level than they were 20 or 25 years ago. However, the cohort growth between the fourth and the eighth grade was the same as, or lower than, it was during the earliest period for which there are NAEP data.

What should be concluded? Is the education system performing better or worse over time?

When the achievement of States was compared, there was little difference in the cohort growth between the fourth and eighth grade. While the State of Maine scored highest in the nation and the State of Arkansas lowest, both States had the same cohort growth, 52 points on the NAEP scale (in mathematics) between the fourth and eighth grade. What should be concluded? Are Maine and Arkansas at the two ends of the school quality continuum, or are they actually equal? 58
A solution: Care needs to be taken when drawing inferences from summary statistics.

Underestimating the negative impact of programs on teacher behaviour

As well as providing useful information for educational decision making, large-scale assessment programs play an important role in communicating values and expectations and have an influence on teacher behaviour. It is dangerous to think that assessment is a ‘neutral’ measuring instrument which only requires further technical developments to make it more effective.59

The problem: Poorly designed assessment systems may provide little support to learning and, at worst, may distort and undermine curriculum intentions, encourage superficial learning, and lower students’ sights on satisfying minimal requirements.60 If sanctions are attached to test results, then teachers typically emphasise what is being tested, thus narrowing and fragmenting the curriculum. The US experience with minimum competency testing (opposite) provides an example of the unintended negative consequences of assessment programs.

A solution: Well-designed assessment systems, which do not focus only on the achievement of minimally acceptable standards, can reinforce curriculum intentions, bringing the intended curriculum and the actual curriculum into closer alignment. They can provide a basis for valuable conversations among teachers about learning and its assessment, and between teachers, students and parents about individuals’ current levels of progress, their strengths and weaknesses, and the kinds of learning experiences likely to be effective in supporting further learning.

Underestimating improvement and accountability tensions

The problem: Developing tests that can be used to hold schools accountable and also to improve instruction may result in a conflict of design and place undue pressure on teachers and schools. Tests used for accountability purposes are usually limited in scope and are thus incapable of providing a comprehensive picture of student achievement. Also if test results are to be used to inform teaching, then they need to be administered early in the year; tests for accountability purposes need to be administered at the end of the year.

If assessment results are used to draw conclusions about the performances of individual teachers or schools, or to allocate resources, then schools may attempt to manipulate data. For example, in Chile, some schools, realising that their rank in the national league tables depended on the reported socio-economic groupings of their students, overestimated the extent of poverty among their students to help boost their position.61

A solution: Clarify the different purposes of systemwide tests and provide teacher professional development to assist teachers to use centrally-collected data to inform teaching. Ensure that the indicators on which decisions are made are incorruptible. Monitor schools’ responses to the assessment program.

Underestimating the negative consequences of accountability measures

The problem: Some evidence suggests that the unintended negative effects of high stakes accountability uses often outweigh the intended positive effects.62 For example, those opposed to the threat of reconstitution of schools in
United States argue that it is a strategy which blames teachers for school failure, demoralising the profession, while doing little to solve the underlying problems that contribute to low performance. (Those in favour of the strategy believe that the threat of reconstitution helps to motivate improvement, particularly in low level or probationary schools. Improvement in these schools is cited as evidence of the positive effect.) Evidence suggests that the impact will be positive or negative depending on the circumstances—which include strong leadership, collective responsibility, a clear break with the past and professional development and capacity building.

A solution: Continue to monitor the unintended consequences of high stakes accountability assessments.

Example 35 US ‘minimum competency’ testing—unintended negative consequences

Minimum competency tests were introduced in the 1970s and 1980s to establish whether students were achieving the minimum levels of knowledge and skill expected of students in particular grades (for example, end of high school). As many commentators have observed, a common response by American teachers to minimum competency tests was to focus their teaching efforts on the foundational skills assessed by these tests and to concentrate their attention on students who had not yet achieved these skills. This was sometimes at the expense of extending the knowledge and skills of higher achieving students. According to some writers, these tests not only constrained classroom teaching, but also had dubious benefits for the students they were designed to serve:

Minimum competency tests are often used as a policy tool to require that students meet some basic level of achievement, usually in reading, writing and computation, with the intention that the use of these tests will lead to the elimination of these educational problems... [However,] the empirical findings show that the negative consequences far outweigh the few positive results... For example, Griffin and Heidorn (1996) showed that minimum competency tests do not help those they are most intended to help—students at the lowest end of the achievement distribution... There have been several studies focusing on the effects of minimum competency testing on curriculum, teaching, and learning. Most of these studies have been critical of their negative effects on curriculum and instruction.63

Other writers are less damning, pointing to evidence of a relationship between minimum competency testing and improved performances among lower-achieving students.64 Nevertheless, because minimum competency tests generally were perceived not to have been effective in raising educational standards, by the 1990s there was a trend away from large-scale tests focused on the achievement of minimally acceptable standards to tests focused on newly valued ‘world class’ standards.65

Example 35 US ‘minimum competency’ testing—unintended negative consequences

Minimum competency tests were introduced in the 1970s and 1980s to establish whether students were achieving the minimum levels of knowledge and skill expected of students in particular grades (for example, end of high school). As many commentators have observed, a common response by American teachers to minimum competency tests was to focus their teaching efforts on the foundational skills assessed by these tests and to concentrate their attention on students who had not yet achieved these skills. This was sometimes at the expense of extending the knowledge and skills of higher achieving students. According to some writers, these tests not only constrained classroom teaching, but also had dubious benefits for the students they were designed to serve:

Minimum competency tests are often used as a policy tool to require that students meet some basic level of achievement, usually in reading, writing and computation, with the intention that the use of these tests will lead to the elimination of these educational problems... [However,] the empirical findings show that the negative consequences far outweigh the few positive results... For example, Griffin and Heidorn (1996) showed that minimum competency tests do not help those they are most intended to help—students at the lowest end of the achievement distribution... There have been several studies focusing on the effects of minimum competency testing on curriculum, teaching, and learning. Most of these studies have been critical of their negative effects on curriculum and instruction.63

Other writers are less damning, pointing to evidence of a relationship between minimum competency testing and improved performances among lower-achieving students.64 Nevertheless, because minimum competency tests generally were perceived not to have been effective in raising educational standards, by the 1990s there was a trend away from large-scale tests focused on the achievement of minimally acceptable standards to tests focused on newly valued ‘world class’ standards.65
Overestimating the strength of league tables

The problem: The evidence suggests that the unintended negative consequences of league tables outweigh any positive contribution they might make. Where league tables do not take students’ intake characteristics into account, schools can be blamed or praised for achievement on the basis of factors beyond the influence of the school.

Using pupil assessment to place schools in rank order, as a way of forcing change in the curriculum or in teaching methods, is recognised as being unfair unless some measure of value-added by the school is used. Even then, ranking may alienate those at the bottom and fail to motivate those at the top of the order; it does not support the message that all schools can improve.

Also, media enthusiasm for the use of league tables often conflicts with the public’s interest in having clearly understandable information. Example 36 illustrates the problem.

A solution: Discourage the use of league tables. If schools are ranked publicly, make the limitations of the data clear.

Underestimating the problems of value added measures

The problem: Value added measures are complex and a single index can give a distorted picture of school performance.

A solution: Use multiple indices of performance and continue to be aware of the limitations of value added information.

Example 36 Media misuse of league tables

In response to the introduction of the Tennessee Value Added Assessment System, Tennessee newspapers printed school league tables even where schools many rankings apart had negligible score differences. The newspaper did not report the evaluators’ clear statement that school scores were unstable and could not be relied on for clear distinctions in performance.

In 1996, one newspaper transformed the value added scores into percentile rankings, even though the technical documentation for the scores did not support the interpretation.
**The problem:** It takes time to collect value added information and by the time information from a particular school has been analysed and provided to the school, the information refers to the achievements of students who entered that school several years previously. Its usefulness for making judgements about school effectiveness for future students may be dubious, especially if there have been staff changes.

**A solution:** Where information is analysed on a yearly basis, make adjustments for prior contributing factors that extend over two or more years in time. Do not judge schools, or teachers within those schools, by the achievements of a single cohort of students, but on their performance over time.

**The problem:** There is a lack of evidence for the positive impact of value added information.

...the enormous national investment in performance data has been something of an act of faith; we need further empirical evidence to help answer such questions as: How do staff in schools actually make use of value added data? Is the effort worthwhile? Do value added measures help to improve education in practice? Under what conditions and with what prerequisites? What kinds of professional development and support are necessary? All of this is crying out to be explored empirically by building up a systematic evidence base.

**A solution:** Monitor the impact of value added information.

**Assuming that summative information will inform teaching**

There is ongoing discussion, particularly in the United Kingdom, about the need to distinguish between formative and summative assessment purposes—formative assessment being assessment for learning (to feed directly into the teaching learning cycle); summative assessment the assessment of learning (for reporting purposes). The assumption that a single assessment can effectively serve both these purposes is contentious.

**The problem:** It is assumed that summative information will inform teaching. However, ‘...there is general agreement that where there are both formative and summative purposes [for assessment], there will invariably be a drift towards more emphasis on the summative functions which inform managerial concerns for accountability and evaluation. The formative functions, which inform teaching and learning, are likely to be diminished.’

For example, in Chile, almost two thirds of teachers reported that they did not use the special manual that dealt with the pedagogical implications of the national test results. (Studies of teacher use of other systemwide achievement data may result in equally sobering findings.)

**A solution:** Provide professional development for teachers to assist them to use achievement data to inform teaching.
Ignoring the role of teachers in reform

In the United Kingdom, much has been written about the 'black box' of the classroom.

In terms of systems engineering, present policy seems to treat the classroom as a black box. Certain inputs from the outside are fed in or make demands—pupils, teachers, other resources, management rules and requirements, parental anxieties, tests with pressures to score highly, and so on. Some outputs follow, hopefully pupils who are more knowledgeable and competent... But what is happening inside?

The problem: The collection of student achievement data will not by itself improve standards. It has been known for a long time that the most effective way of improving the quality of education for individual pupils is for teachers in schools to evaluate what they are doing and to make the necessary changes.

A solution: Provide professional development for teachers to assist them to use achievement data to inform teaching.

A CASE STUDY

The US National Education Goals Panel reports progress on 33 indicators linked to eight National Education Goals. In the 1997 report two States—North Carolina and Texas—stood out for realising positive gains on the greatest number of indicators. An analysis of the reforms in both States was undertaken to identify the factors that could and could not account for their progress. The findings of the study are summarised here.

Factors commonly associated with student achievement which did not explain the test score gains included real per pupil spending, teacher/pupil ratios, the number of teachers with advanced degrees, and the experience level of teachers in the system.

Two plausible explanations for test score gains were proposed: the way in which policies were developed, implemented and sustained (the policy environment); and the policies themselves.

The policy environment

- Leadership from the business community

In both States the business community played a critical leadership role in developing and sustaining reform including funding organisations that brought together the business, education and policy-making communities. Business involvement was also characterised by the presence of a few business leaders who became deeply involved.

- Political leadership

Political leadership was essential at critical points in the reform process. The passage of legislation involved coalitions from both parties, and the business community remained a consistent external voice.

- Consistency of the reform agenda

Despite changes in Governors and legislators the reform agenda has been maintained.
The policies

- Systemwide standards by grade for clear teaching objectives
- Holding all students to the same standards

The greatest improvement in mathematics and reading scores is among African American, Hispanic and economically disadvantaged children.

- Systemwide assessment closely linked to academic standards
- Accountability systems with consequences for results

Both States rate schools based on their performance on State tests, provide monetary rewards for schools based on their test performance, and have the power to disenfranchise schools and remove principals based on sustained levels of poor performance. In rating schools, both States adjust beginning gain scores to the actual students in the schools at the beginning of the year to take account of schools with high student turnover or large numbers of new students with low or high test scores. Both also keep close scrutiny of the students excluded from taking the tests to protect against manipulation by teachers or principals.

- Increasing local control and flexibility for administrators and teachers
- Computerised feedback systems, data for continuous improvement

Scores on the tests are provided to students, parents, teachers, schools and school districts. Access to school level results is provided on the Internet in both States, and both States have developed varied formats for reporting test results by sets of questions related to key learning objectives. Teachers in both States have access to summaries and individual tests of students entering their classes each year.

- Shifting resources to schools with more disadvantaged students

Both States have gradually shifted resources to schools with more disadvantaged students. The sustainability of the above policies may rely on the distribution of resources among schools and districts remaining fair and equitable.

Both States have built a substantial infrastructure for supporting a process of continual improvement in education which involves the whole community. A separate recent study of 26 high achieving, high poverty schools in Texas exhibited the following characteristics: a strong focus on ensuring academic success for each student, a refusal to accept excuses for poor performance, a willingness to experiment with a variety of strategies, intensive and sustained efforts to involve parents and the community, an environment of mutual respect and collaboration, and a passion for continuous improvement and professional growth.
A CHECKLIST OF CONSIDERATIONS

I believe that there is much that is wrong with this system, that there are signs here and there of improvement, and that there are ways to make assessment much better in serving teaching and learning. We have more and more of these numbers, but they are too often not adding up to good information.78

Achievement data will be most useful to policy makers if they provide evidence for informed decision making. Data on which improvement and accountability decisions are made need to be valid and reliable and the impact of assessment and reporting initiatives needs to be closely monitored.

The final section of this guide provides a checklist of considerations for ensuring that systemwide achievement data provide evidence for informed decision-making that will lead to improved student learning.

Data collection

1 Is the purpose of the program clear?

The purpose for which data are collected must be clear and the data collection procedures appropriate for the purpose.

For example, if data are to be used to target resources at system level, then information identifying population subgroups needs to be collected. If data are to be used to track changes in achievement over time, then procedures need to be put in place to link each new study to a previous study. If the program is to provide parents of all students at a particular year level with a report, then the program needs to be full cohort rather than representative sample.

2 Will the data be valid?

The kinds of data collected must be appropriate for the purpose of the program.

For example, if the program is to provide results which can be used by teachers to improve student learning, then tests need to address clearly defined learning outcomes to which teachers teach. The tests need to address a wide range of outcomes and they need to be instructionally sensitive. That is, effective instruction will produce improvements in performance.

If the data are to be used to provide summary statistics on the achievements of all students and data on the achievements of subgroups of the population, then tasks need to be fair and meaningful to all students. That is, the tasks must allow students of different gender and from different cultural, social and economic backgrounds to demonstrate what they know and can do.

3 Will the data be reliable?

The information needs to be accurate.

For example, the results need to allow accurate generalisation about student achievement. (Assessment tasks must address an adequate range of learning outcomes.) Procedures must be in place to collect comparable achievement data, (trial testing of all instruments, uniform administration, marking, and recording procedures). If the program is based on the collection of data from a representative sample, the sampling must be conducted so that the standard errors are acceptable in relation to the policy decisions that will be based on results.
4 Have the known negative consequences been considered?
Attempts must be made to avoid known unintended negative consequences of assessment programs.
For example, given what is known about the effects of minimum competency testing, systemwide tests should be designed to provide information about both low level and higher order skills.

Reporting
5 Will reports be comprehensive?
Analyses need to be described clearly and arguments about the interpretation of analyses reported. Measurement uncertainty needs to be shown.

6 Will reporting be accessible to different stakeholders?
Different kinds of reports (containing different kinds of information) need to be provided for stakeholders at different levels of the education system.
For example, reports for teachers and schools must provide information that can impact on instruction.

Monitoring Trends
7 Will reliable comparisons over time be facilitated?
When interpreting comparisons, like needs to be compared with like. School and student-level exclusions need to be detailed and the impact of these on comparisons of means and distributions assessed. The reliability of different kinds of statistics for monitoring purposes needs to be considered in relation to the kind of information provided.

For example, when monitoring trends over time, the percentage of students achieving a particular performance standard is in general a less reliable statistic than the group mean.

Accountability measures
8 Publication of data
If achievement data are to be made publicly available, then the limitations of the data need to be made clear. The impact of publication needs to be monitored closely, especially the impact of school league tables.

9 Will rewards and sanctions be accompanied by capacity building?
Providing rewards and sanctions alone will not improve teaching and learning—professional development programs need to be in place.

Monitoring assessment programs
10 Are there procedures in place to monitor the usefulness of study findings?
As part of the commitment to collecting systemwide achievement data, research to monitor the impact and usefulness of findings should be undertaken, especially the monitoring of more recent initiatives such as value-added reporting.
USEFUL WEBSITES FOR POLICY MAKERS

**United States**

**Council of Chief State School Officers (CCSSO)**
A nation-wide, non-profit organisation composed of the public officials who head departments of elementary and secondary education in the United States. CCSSO seeks members' consensus on major educational issues and expresses their view to civic and professional organisations, federal agencies, Congress and the public.

[www.cccso.org](http://www.cccso.org)

**Consortium on Chicago School Research**
An independent federation of Chicago area organisations that conducts research on ways to improve Chicago's public schools and ways to assess the progress of school improvement and reform.

[www.consortium-chicago.org](http://www.consortium-chicago.org)

**National Education Goals Panel**
A bipartisan and intergovernmental body of federal and state officials created in July 1990 to assess and report state and national progress toward achieving the National Education Goals. The NEGP tracks and annually reports on 33 indicators linked to the eight National Education Goals to provide a picture of individual state success in improving education.

[www.negp.gov](http://www.negp.gov)

**US Regional Educational Laboratories**
A network of 10 regional laboratories administered by the Office of Educational Research and Improvement (OERI) working with schools and school districts, states, research institutions, and community based organisations to explore effective ways to implement educational reforms. Each lab has been asked to develop expertise in one specialty area.

Northeast and Islands Laboratory at Brown University (LAB) specialty area: language and cultural diversity
[www.lab.brown.edu](http://www.lab.brown.edu)

Mid Atlantic Laboratory for Student Success (LSS) specialty area: urban education
[www.temple.edu/LSS](http://www.temple.edu/LSS)

Appalachia Educational Laboratory (AEL) specialty area: rural education
[www.ael.org](http://www.ael.org)

The SouthEastern Regional Vision for Education (SERVE) specialty area: early childhood education
[www.serve.org](http://www.serve.org)
<table>
<thead>
<tr>
<th>Region</th>
<th>Specialty Area</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Central Regional</td>
<td>Technology</td>
<td><a href="http://www.ncrel.org">www.ncrel.org</a></td>
</tr>
<tr>
<td>Regional Laboratory (NCREL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southwest Educational</td>
<td>Language and cultural diversity</td>
<td><a href="http://www.sedl.org">www.sedl.org</a></td>
</tr>
<tr>
<td>Development Laboratory (SEDL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid-continent Regional</td>
<td>Curriculum, learning, and instruction</td>
<td><a href="http://www.mcrel.org">www.mcrel.org</a></td>
</tr>
<tr>
<td>Educational Laboratory (McREL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WestEd</td>
<td>Assessment and accountability</td>
<td><a href="http://www.WestEd.org">www.WestEd.org</a></td>
</tr>
<tr>
<td>Northwest Regional Educational</td>
<td>School change processes</td>
<td>[www nwrel.org](<a href="http://www">http://www</a> nwrel.org)</td>
</tr>
<tr>
<td>Laboratory (NWREL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pacific Resources for Education and Learning (PREL)</td>
<td>Language and cultural diversity</td>
<td><a href="http://www.prel.org">www.prel.org</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
End Notes


8. Connecticut Department of Education www.state.ct.us/sde


12. Education Department of Western Australia (1998) Student Achievement in Mathematics. Perth: Education Department of Western Australia.


18. Pennsylvania Department of Education www.pde.psu.edu/pssa/stand

19. Australian Council for Educational Research

20. Department of Education and Children’s Services, South Australia Putting the basic skills test results to work in the classroom.


23. Department for Education and Children’s Services, South Australia. See end note 20.


26 Connecticut Department of Education. See end note 8.


29 Department for Education and Children’s Services, South Australia. See end note 20.


37 California Department of Education www.cde.ca.gov


39 Maryland Department of Education http://www.msde.state.md.us


41 Australian Council for Educational Research


46 Kentucky Department of Education www.kde.state.ky.us

47 Colorado Department of Education www.ced.state.co.us/

48 North Carolina Public Schools infoweb www.dpi.state.nc.us/abc_results/results_99/exe_summary


Colorado at Boulder.
http://cresst96.cse.ucla.edu/CRESST/pages/reports.htm


