Benchmarking Learning Systems: Student Performance Assessment in International Context

Linda Darling-Hammond
with the assistance of Laura Wentworth

Stanford University

scope
Stanford Center for
Opportunity Policy in Education

This study was conducted by the Stanford Center for Opportunity Policy in Education (SCOPE) with support from the Ford Foundation and the Nellie Mae Education Foundation.

The Stanford Center for Opportunity Policy in Education (SCOPE) supports cross-disciplinary research, policy analysis, and practice that address issues of educational opportunity, access, equity, and diversity in the United States and internationally.

Citation: Darling-Hammond, L & Wentworth, L. (2010). *Benchmarking learning systems: Student performance assessment in international context*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Stanford Center for Opportunity Policy in Education
Barnum Center, 505 Lasuen Mall
Stanford, California 94305
Phone: 650.725.8600
scope@stanford.edu
http://edpolicy.stanford.edu

scope
Stanford Center for
Opportunity Policy in Education

Abstract

High-performing nations integrate curriculum, instruction, and assessment to improve both teaching and learning. As a large and increasing part of their examination systems, they use open-ended performance tasks and school-based assessments to give students opportunities to develop 21st century skills: The abilities to find and organize information to solve problems, frame and conduct investigations, analyze and synthesize data, and apply learning to new situations. This paper illustrates how several nations integrate these assessments into the curriculum to create stronger learning for both students and teachers, resulting in higher and more equitable achievement.

Preface and Acknowledgements

This paper is one of eight written through a Stanford University project aimed at summarizing research and lessons learned regarding the development, implementation, consequences, and costs of performance assessments. The project was led by Linda Darling-Hammond, Charles E. Ducommun Professor of Education at Stanford University, with assistance from Frank Adamson and Susan Shultz at Stanford. It was funded by the Ford Foundation and the Nellie Mae Education Foundation and guided by an advisory board of education researchers, practitioners, and policy analysts, ably chaired by Richard Shavelson, one of the nation's leading experts on performance assessment. The board shaped the specifications for commissioned papers and reviewed these papers upon their completion. Members of the advisory board include:

Eva Baker, Professor, UCLA, and Director of the Center for Research on Evaluation, Standards, and Student Testing

Christopher Cross, Chairman, Cross & Joftus, LLC

Nicholas Donahue, President and CEO, Nellie Mae Education Foundation, and former State Superintendent, New Hampshire

Michael Feuer, Executive Director, Division of Behavioral and Social Sciences and Education in the National Research Council (NRC) of the National Academies

Edward Haertel, Jacks Family Professor of Education, Stanford University

Jack Jennings, President and CEO, Center on Education Policy

Peter McWalters, Strategic Initiative Director, Education Workforce, Council of Chief States School Officers (CCSSO) and former State Superintendent, Rhode Island

Richard Shavelson, Margaret Jacks Professor of Education and Psychology, Stanford University

Lorrie Shepard, Dean, School of Education, University of Colorado at Boulder

Guillermo Solano-Flores, Professor of Education, University of Colorado at Boulder

Brenda Welburn, Executive Director, National Association of State Boards of Education

Gene Wilhoit, Executive Director, Council of Chief States School Officers

The papers listed below examine experiences with and lessons from large-scale performance assessment in the United States and abroad, including technical advances, feasibility issues, policy implications, uses with English language learners, and costs.

- Jamal Abedi, *Performance Assessments for English Language Learners.*
- Linda Darling-Hammond, with Laura Wentworth, *Benchmarking Learning Systems: Student Performance Assessment in International Context.*
- Suzanne Lane, *Performance Assessment: The State of the Art.*
- Raymond Pecheone and Stuart Kahl, *Developing Performance Assessments: Lessons from the United States.*
- Lawrence Picus, Frank Adamson, Will Montague, and Maggie Owens, *A New Conceptual Framework for Analyzing the Costs of Performance Assessment.*
- Brian Stecher, *Performance Assessment in an Era of Standards-Based Educational Accountability.*
- Barry Topol, John Olson, and Edward Roeber, *The Cost of New Higher Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments.*

An overview of all these papers has also been written and is available in electronic and print format:

- Linda Darling-Hammond and Frank Adamson, *Beyond Basic skills: The Role of Performance Assessment in Achieving 21st Century Standards of Learning.*

All reports can be downloaded from http://edpolicy.stanford.edu.

We are grateful to the funders, the Advisory Board, and these authors for their careful analyses and wisdom. These papers were ably ushered into production by Barbara McKenna. Without their efforts, this project would not have come to come to fruition.

Benchmarking Learning Systems: Student Performance Assessment in International Context

Since the release of *A Nation at Risk* in 1983, the United States has launched a set of wide-ranging reforms to better prepare all children for the higher educational demands of life and work in the 21st century. All 50 states have developed standards for learning and tests to evaluate student progress. No Child Left Behind reinforced the use of test-based accountability to raise achievement, and scores have climbed on state tests used for accountability purposes, yet the United States has fallen further behind on international assessments of student learning since the law was passed in 2001.

On the Program in International Student Assessment (PISA) tests in 2006, the United States ranked 25th of 30 OECD countries in mathematics and 21st of 30 in science, a decline in both raw scores and rankings from three years earlier. (Reading scores were not reported, because of editing problems with the U.S. test.) Furthermore, U.S. students scored lowest on the problem-solving items. The United States also had a much wider achievement gap than the highest-ranked jurisdictions, such as Finland, Canada, Australia, New Zealand, Hong Kong, Korea, and Japan. Singapore, which has not yet received results from PISA, places at the top of the rankings on the International Education Assessments (TIMSS and PIRLS), where these other nations also excel. It also shows a very narrow distribution of achievement, despite its multiethnic, multilingual, socioeconomically diverse population.

It is worth noting that PISA assessments focus explicitly on 21st century skills, going beyond the question posed by most U.S. standardized tests, "Did students learn what we taught them?" to ask, "What can students do with what they have learned?" (Stage, 2005). PISA defines literacy in mathematics, science, and reading as the ability to *apply* knowledge to new problems and situations. This kind of higher-order learning is increasingly emphasized in other nations' assessment systems, but often discouraged by the multiple-choice tests most U.S. states have adopted.

Policy discussions in Washington often refer to these international rankings when emphasizing the need to create more "internationally competitive" standards by benchmarking expectations in the United States to those in high-performing nations. Typically, the focus is on identifying topics taught at each grade level in these countries. The analyses reveal that higher-achieving countries teach fewer topics more deeply each year, focus on applications of knowledge, rather than recall of facts, and have a more thoughtful sequence of expectations based on developmental learning progressions within and across domains (Schmidt, Wang, & McKnight, 2005; Valverde & Schmidt, 2000).

It is also important to examine *how* these topics are taught and assessed—so that we understand how the top education systems shape what students actually learn and can do. European and Asian nations where student learning has improved dramatically have created curriculum guidance and assessments focused explicitly on 21st-century skills: the abilities to find and organize information to solve problems, frame and conduct investigations, analyze and synthesize data, apply learning to new situations, self-monitor and improve one's own learning and performance, communicate well in multiple forms, work in teams, and learn independently.

The forms of testing used in the United States and those used in higher-achieving countries reinforce the sharp divergences between curricula. Whereas U.S. tests rely primarily on multiple-choice items that evaluate recall and recognition of discrete facts, most high-achieving countries primarily rely on open-ended items that require students to analyze, apply knowledge, and write extensively. Furthermore, these nations' growing emphasis on project-based, inquiry-oriented learning has prompted increased use of school-based tasks, which include research projects, science investigations, development of products, and related reports or presentations. These assessments, which are incorporated into the overall examination scoring system, help focus the day-to-day work of teaching and learning on the development of higher-order skills and use of knowledge to solve problems.

Smaller countries often have a system of national standards that are sometimes—though not always—accompanied by national tests in the upper grades. Top-ranking Finland uses local assessments almost exclusively to evaluate its national standards and manages a voluntary national assessment for college admissions at only one grade level. Larger nations—such as Canada, Australia, and China—have state- or provincial-level standards, and their assessment systems are typically a blend of state and local assessments. Managing assessment at the state rather than national level, where it remains relatively close to the schools, turns out to be an important way of enabling strong teacher participation and ensuring high-quality local assessments that can be moderated to ensure consistency in scoring.

In many cases, school-based assessments complement centralized "on-demand" tests and may constitute up to 60% of the final examination score. Tasks are mapped to the standards or syllabus for the subject and are selected because they represent critical skills, topics, and concepts. They are often outlined in the curriculum guide, but they are generally designed, administered, and scored locally, based on common specifications and evaluation criteria. Whether locally or centrally developed, decisions about when to undertake these tasks are made at the classroom level, so they are used when appropriate for students' learning process. Teachers can get information and provide feedback as needed, something that traditional standardized tests do not enable them to do. In addition, as teachers use and evaluate these tasks, they become more knowledgeable about both the standards and how to teach to them. They also become more aware of their students' learning needs. Thus, the process improves the quality of teaching and learning.

Like the behind-the-wheel test given for all new drivers, these performance assessments evaluate what students can actually *do,* not just what they know. The road test not only reveals some important things about drivers' skills, preparation for the test helps *improve* those skills as novice drivers practice to get better. In the same way, performance assessments set a standard toward which everyone must work. The task and the standards are not secret, so teachers and students know what skills they need to develop and how they will need to be demonstrated.

Finally, these countries typically do not use their examination systems to punish schools or to deny diplomas to students. Following the problems that resulted from the Thatcher government's use of test-based school rankings, which caused a narrowing of the curriculum and widespread exclusions of low-performing students from school (Rustique-Forrester, 2005), several countries enacted legislation precluding the use of test results for school rankings. Instead, high school examinations provide information for higher education, vocational training, and employment. Students often choose the areas in which they will be examined to demonstrate their qualifications. The systems are focused on using information for users of the system and for curriculum improvement, rather than sanctions. Thus, governments can set higher standards and work with schools to achieve them, rather than devising tests and setting cut scores at a minimal level to avoid dysfunctional side-effects.

In this paper, we examine the assessment systems of several high-achieving education systems: two Scandinavian nations—Finland and Sweden—plus a group of English-speaking jurisdictions that have some shared approaches to assessment, as well as some interesting variations—Australia, Hong Kong, Singapore, and the United Kingdom. In addition, we describe the International Baccalaureate system, which is now used in many U.S. schools as well as schools around the world. (A summary of assessment system features is included in Appendix A). These examples provide interesting lessons about how assessments can be linked to curriculum and integrated into the instructional process to shape and improve learning for students and teachers alike.

Finland

Finland has been a poster child for school improvement since it rapidly climbed to the top of the international rankings after it emerged from the Soviet Union's shadow. It now ranks first among all the OECD nations on the PISA assessments in mathematics, science, and reading. Leaders in Finland attribute these gains to their intensive investments in teacher education and major overhaul of the curriculum and assessment system. Prospective teachers are competitively selected from the pool of college graduates and enter a three-year graduate-level teacher preparation program, which is entirely free of charge and comes with a living stipend. The master's degree program includes both extensive coursework on how to teach—with a dual focus on inquiry-oriented teaching and teaching that meets the needs of diverse learners—plus at least a full year of clinical experience in a school associated with the university. Preparation includes a strong

focus on how to use formative performance assessments in the service of student learning (Laukkanen, 2008; Buchberger & Buchberger, 2004.).

Policy makers decided that if they invested in very skillful teachers, they could allow local schools more autonomy to decide what and how to teach—a reaction against the highly centralized system they sought to overhaul. Finland's national core curriculum is now a much leaner document, reduced from hundreds of pages of highly specific prescriptions to descriptions of a small number of skills and core concepts. (For example, about 10 pages describe the full set of math standards for all grades.) This guides teachers in collectively developing local curricula and assessments that encourage students to be active learners who can find, analyze, and use information to solve problems in novel situations.

Finland has no external standardized tests used to rank students or schools. Although it may seem counterintuitive to Americans accustomed to external testing as a means of accountability, Finland's leaders point to its use of school-based, student-centered, open-ended tasks embedded in the curriculum as an important reason for the nation's extraordinary success on international exams (Lavonen, 2008; Finnish National Board of Education, 2007). School-level samples of student performance are evaluated periodically by the Finnish education authorities, generally at the end of the 2nd and 9th grades, to inform curriculum and school investments. All other assessments are designed and managed locally. The national core curriculum provides teachers with recommended assessment criteria for specific grades in each subject and in the overall final assessment of student progress each year (Finnish National Board of Education, June 2008). Local schools and teacher then use those guidelines to craft a more detailed curriculum and set of learning outcomes at each school as well as approaches to assessing benchmarks in the curriculum (Finnish National Board of Education, June 2008). Teachers are treated as "pedagogical experts" who have extensive decision-making authority in the areas of curriculum and assessment in addition to other areas of school policy and management (Finnish National Board of Education, April 2008).

According to the Finnish National Board of Education (June 2008), the main purpose of assessing students is to guide and encourage students' own reflection and self-assessment. Consequently, ongoing feedback from the teacher is very important. Teachers give students formative and summative reports both through verbal feedback and on a numerical scale based on students' level of performance in relation to the objectives of the curriculum. All Finnish schools use a grading scale of 4-10, where 5 is "adequate" and 10 is "excellent." The recommended assessment criteria are shaped around the grade of 8 or "good." Teachers' reports must be based on multiple forms of assessment, not only exams. Schools are responsible for giving basic education certificates for completing the different milestones of comprehensive school up to ninth grade and additional classes prior to university (European Commission, 2007/2008).

Most Finnish students take a set of voluntary matriculation examinations that pro-

vide information for university admissions based on students' abilities to apply problem solving, analytic, and writing skills. University and high school faculty members construct the examinations—which are composed of open-ended essays and problem solutions—under the guidance of the Matriculation Exam Board, which is appointed by the Finnish Ministry of Education to organize, manage, and administer the exam (The Finnish Matriculation Examination, 2008). The Board members (about 40 in number) are faculty and curriculum experts in the subject areas tested, nominated by universities and the National Board of Education. More than 300 associate members—also typically high school and college faculty—help develop and review the tests. High school teachers grade the matriculation exams locally using official guidelines, and samples of the grades are re-examined by professional raters hired by the Board (Kaftandjieva & Takala, 2002).

Students take at least four exams, with the test in the students' mother tongue (Finnish, Swedish, or Saami) being compulsory. These tests have a textual skills section, which evaluates students' analytic skills and linguistic expression, and an essay that focuses on the development of thinking, linguistic expression, and coherency. They then choose three other tests from among the following: the test in the second national language, a foreign language test, the mathematics test, and one or more tests from the general studies battery of tests in the sciences and humanities (e.g. religion, ethics, philosophy, psychology, history, social studies, physics, chemistry, biology, geography, and health education). The tests also incorporate questions which cross disciplinary boundaries.

It is interesting to note that this system assumes that all students aiming for college (who comprise a majority of Finnish students) will be at least bilingual, and many will be trilingual. The language tests evaluate listening and reading comprehension as well as writing.

In addition to the choice of which tests to take, students choose which items to answer within the exams. In the general battery, they are typically given a set of questions or prompts from which they must respond to six or eight of their choice. The mathematics test has about15 problems from which they must choose 10 to answer.

Problems require critical thinking and modeling, as well as straightforward problem-solving. For example, the basic mathematics exam poses this problem:

> A solution of salt and water contains 25% salt. Diluted solutions are obtained by adding water. How much water must be added to one kilogram of the original solution in order to obtain a 10% solution? Work out a graphic representation which gives the amount of water to be added in order to get a solution with 2-25% of salt. The amount of water (in kilograms) to be added to one kilogram of the original solution must be on the horizontal axis; the salt content of the new solution as a percentage must be on the vertical axis.

And the advanced mathematics exam poses this one:

> In a society the growth of the standard of living is inversely proportional to the standard of living already gained, i.e. the higher the standard of living is, the less willingness there is to raise it further. Form a differential-equation-based model describing the standard of living and solve it. Does the standard of living rise forever? Is the rate of change increasing or decreasing? Does the standard of living approach some constant level?

Assessment is used in Finland to cultivate students' active learning skills by posing complex problems and helping students address these problems. For example:

> In a Finnish classroom, it is rare to see a teacher standing at the front of a classroom lecturing students for 50 minutes. Instead, students are likely to determine their own weekly targets with their teachers in specific subject areas and choose the tasks they will work on at their own pace. A description of a Finnish school (Korpela, 2004) illustrates how students are likely to be walking around, rotating through workshops or gathering information, asking questions of their teacher, and working with other students in small groups. They may be completing independent or group projects or writing articles for their own magazine. The cultivation of independence and active learning allows students to focus on broad knowledge with emphasis on analytical thinking, problem solving, and metacognitive skills. These types of skills are increasingly emphasized on tests such as the PISA, which seek to evaluate students' capacities to think independently and creatively in applying knowledge (Lavonen, 2008).

Sweden

Over the past 40 years, Sweden's national assessment system has, like Finland's, shifted from a centralized system based on one test to a more localized system based on multiple forms of assessments. With this change, Sweden hoped to increase Upper Secondary school enrollment and provide more open-access to higher education. Around 1970, Sweden abolished its *studentexamen*, a nationally administered exit exam that ranked Upper Secondary students and placed them in higher education programs (European Commission, 2006-2007). The new policy was intended to produce more valid and reliable evidence of whether students would succeed in university. Additionally, the country wanted to correct social and educational inequities caused by a one-size-fits-all assessment system (Eckstein and Noah, 1993).

Sweden's national curriculum is adjusted for the local context. In Compulsory school, ages 7 to 16, the common curriculum, which includes nationally approved syllabi for individual subjects, is adapted in each district to address local conditions (Swedish National Agency for Education, 2005). In Upper Secondary school, ages 17 to 20,

Sweden has 17 subject matter curricula consisting of three-year programs providing a general education and eligibility to study at the post-secondary or university level. The programs include eight core subjects (English, the arts, physical education and health, mathematics, general science, social studies, Swedish or Swedish as a second language, and religion) and a set of subject specific area such as "the construction program" or "the business program" that combine general courses with specialized classes. The National Agency for Education determines the required courses for a national specialization and most of the programs require at least 15 weeks of workplace training outside of school.

Sweden pairs its nationally outlined and locally implemented curriculum with multiple layers of assessment controlled by schools and teachers. Assessments in compulsory school consist of several components. First, during each school term, the teacher, student, and parent meet to discuss the student's learning and social development (Swedish National Agency for Education, 2005). Second, students receive grades from their teachers in each term of year 8 (age 15) and the end of the fall term of year 9 (age 16). Teachers base their year 9 grades on the goals in the syllabi, awarding grades of "pass", "pass with distinction," or "pass with special distinction" based on nationally approved assessment criteria (Swedish National Agency for Education, 2005).

Third, schools can use a number of optional diagnostic materials to assess students in Swedish, Swedish as a second language, mathematics, and English. The diagnostic materials help teachers assess students and support their learning. The diagnostic materials in years 6 through 9 assess where students stand in relation to the goals set by the syllabi (Swedish National Agency for Education, 2005).

Fourth, students take nationally approved examinations in year 9. The exams assess the subjects of Swedish, Swedish as a second language, English, and mathematics. Teachers use these assessments as one factor in determining students' grades. The exam at year 9 is compulsory for schools, but not for students. Sweden uses the scores from the test to ensure the grades given by teachers compare to the national standards (Qualifications and Curriculum Authority, 2008). In some jurisdictions, schools give an examination at the end of year 5 in these same subjects.

Towards the end of their Upper Secondary schooling, Swedish students receive a grade in each course and a final grade or "learning certificate" that acts as a compilation of all grades awarded for courses and projects completed by the student. Teachers keep extensive records of student progress, using three assessments to aid in their grading at the Upper Secondary school level: 1) coursework, 2) assessments designed by teachers based on the course syllabi, and 3) nationally approved examinations when grading the core subjects of Swedish, English and mathematics, and selected other areas (Swedish National Agency for Education, 2005). Regional education officials and schools provide time for teachers to calibrate their grading practices to minimize variation across the schools and across the region (Eckstein & Noah, 1993).

While university personnel produce the national exams at year 9 and at the Upper Secondary level, teachers help design the tasks and questions and grade the assessments (Eckstein & Noah, 1993; O'Donnell, 2004). Teachers also design school-based assessments based on the nationally determined syllabi, which dictate the content of the coursework students complete in their classes (Qualifications and Curriculum Authority, 2008a; O'Donnell, 2004; Eckstein & Noah, 1993).

The National School Board examinations administered during Compulsory and Upper Secondary schooling use an open-ended, authentic approach to assessing students. The exam questions are grounded in real world contexts, asking students to use analytic skills and draw on content knowledge learned during their classes. For example, Sweden's native language test at the upper elementary school level asks students about a broad theme. One year, the exam used the theme of "travel" and provided students with contemporary poem, prose, and poetry extracts from a variety of authors, a practical description of how to plan a trip, and data about travel presented in a set of texts, charts, and statistical tables. Schools gave students materials a week in advance, so students had time to review the materials. Students had five hours to write an essay on the topic of their choice that was evaluated on specific criteria emphasized in the syllabus from their course. The skills assessed included using appropriate language in certain circumstances, comprehending the different purposes of language, persuasive mechanisms, presenting information, as well as creative self expression, word choice, and grammar (Eckstein & Noah, 1993, p. 119).

These examples from math assessments illustrate how questions are embedded in real-world contexts. Here is an on-demand sample question from the grade 5 exam that asks students (ages 11 to 12) to grapple with a problem that they might have in their own lives, both weighing and balancing decisions as well as applying math knowledge:

> Carl bikes home from school at four o'clock. It takes about a quarter of an hour. In the evening he's going back to school because the class is having a party. The party starts at 6 o'clock. Before the class party starts, Carl has to eat dinner. When he comes home, his grandmother calls, who is also his neighbor. She wants him to bring in her post before he bikes over to the class party. She also wants him to take her dog for a walk, then to come in and have a chat. What does Carl have time to do before the party begins? Write and describe below how you have reasoned (Petterson, 2008).

The mathematics exam from the Upper Secondary level also frames the questions in real world, tangible topics, and formats. Students have almost four hours to answer 15 questions. The first 10 questions require short answers and the last five questions require longer answers for which students show their work.

Swedish Mathematics Exam at the Third Year of Upper Secondary

Short Format Questions
A coffee blender mixed *x* kg of coffee costing *a* knoner/kg with *y* kg of coffee costing *b* kroner/kg. Give a formula for the price per kg of the blend.
In 1976 Lena had a monthly salary of 6,000 kr. By 1984 her salary had risen to 9,000 kr. In current prices, her salary had risen by 50%. How large was the percent change in fixed prices? In 1976 the Consumer Price Index (CPI) was 382; in 1984 it was 818.
Long Format Questions
A business paid into a pension fund at the beginning of every year a sum of 15,000 kr. The fund has a yearly growth rate of 10%. The first payment was made in 1987 and the last will be in 2010. The pension fund will continue to grow until 2015. How much more will the business have in the fund at the beginning of 2015, if it pays in the same amount as above, but the rate of growth is 15%?

Source: Eckstein & Noah, 1993, pp. 270-272

Australia

In contrast to smaller countries such as Finland and Sweden, which have national curriculum guidance, in Australia each state has its own curriculum and assessment program. In most states, local school-based performance assessment is a well-developed part of the system. In some cases, states have also developed centralized assessment with performance components. Two of the highest-achieving states, Queensland and A.C.T., have highly developed systems of local performance assessment. Victoria, which uses a blended model of centralized and school-based assessment, also generally performs well on national and international tests. An effort to develop common national curriculum standards is underway, but states will continue to be responsible for translating these standards into an assessment program at the state and local levels.

Queensland

In Queensland, there has been no assessment system external to schools for 40 years. Until the early 1970s, a traditional "post-colonial" examination system controlled the curriculum. When it was eliminated, all assessments became school-based. Teachers develop, administer, and score the assessments in relation to the national curriculum guidelines and state syllabi (also developed by teachers), and panels that include teachers from other schools as well as at least one professor from the tertiary education system moderate the assessments.

To create the standards used throughout the province, the central authority gathers groups of teachers and subject experts to write standards that specify different levels of achievement and describe the characteristics of student work at each level. In the excerpt from Queensland's science standards shown below, the left side describes the objectives or "Essential Learnings" that teachers must teach and assess. The objectives convey the knowledge or skill expected at each standard. The standard descriptors to the right detail the expected characteristics and quality of the work. The teachers and

experts also develop samples of work used as exemplars for the different levels. These standards guide the assessments teachers develop and their scoring.

The syllabi seek to strike a balance between "informed prescription" and "informed pro-

Excerpt from Queensland Science Standards

What must be taught and assessed

Expected quality of the work

Investigative processes	The Standards descriptors	
Students need to ... investigate scientifically ... judge the worth of quantitative and qualitative data ... interpret ... apply the outcomes. Students manipulate and review data and scientific techniques ... improve their scientific knowledge. They synthesise research that they have generated to discuss the outcomes in relation to their initial purpose. By the end of the course, students should be able to:	**Standard A**	**Standard C**
	The student work has the following characteristics:	The student work has the following characteristics:
• **conduct and appraise Physics research tasks** — this includes the abilities to formulate questions, hypothesise, plan, manage, evaluate, refine and justify decisions made during investigations, as well as the critical reflection required to fulfil research goals	• formulation of **justified significant** questions/ hypotheses which inform **effective** and **efficient** design, refinement and management of investigations	• formulation of questions and hypotheses to select and manage investigations
• **operate scientific equipment and technology safely** — this includes the abilities to safely select, adapt and apply technologies, laboratory and fieldwork equipment, and consider its limitations; it also incorporates the ability to do this individually and in groups	• assessment of risk, **safe** selection and adaptation of equipment, and appropriate application of technology to gather, record and **process valid** data	• assessment of risk, safe selection of equipment, and appropriate application of technology to gather and record data
• **use primary and secondary data** — this includes the abilities to analyse and extrapolate from data, and to identify relationships, patterns and anomalies in primary and secondary data.	• **systematic** analysis of primary and secondary data to **identify relationships** between patterns, trends, errors and anomalies.	• analysis of primary and secondary data to identify obvious patterns, trends, errors and anomalies.

fessionalism." They spell out a small number of key concepts and/or skills to be learned in each course, and what kinds of projects or activities (including minimum assessment requirements) students should be engaged in. Each school designs its program to fit the needs and experiences of its own students, choosing specific texts and topics with this in mind. However, all schools must use shared, specific criteria for evaluation of student work, based on the course objectives and specific standards for an A, B, C, D, and E.

As the criteria from the physics syllabus (page 13) indicate in substantial detail, in the category of *Knowledge and conceptual understanding*, work that meets an "A" standard demonstrates interpretation, comparison, and explanation of complex concepts, theories and principles, whereas work at an "E" standard is characterized by reproduction of isolated facts and application of simple, given algorithms. In this particular course, objectives also include *Investigative processes,* and *Evaluating and concluding,* with indicators spelled out for all of these objectives. The expectations of work quality are challenging, as shown in the example on page 11.

Extended Experimental Investigations

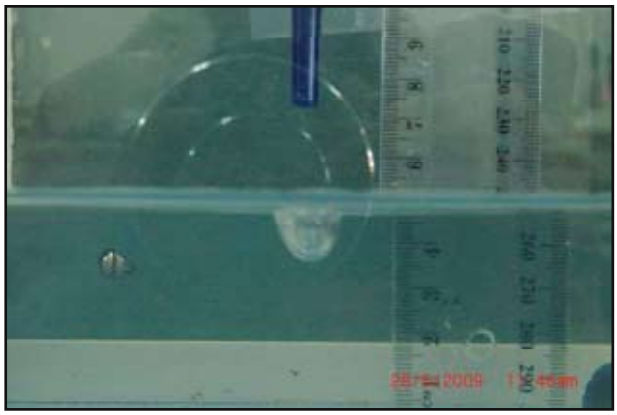
In Queensland science courses, students must complete an extended experimental investigation. The instructions for the task read:

> Within this category, instruments are developed to investigate a hypothesis or to answer a practical research question. The focus is on planning the extended experimental investigation, problem solving and analysis of primary data generated through experimentation by the student. Experiments may be laboratory or field based. An extended experimental investigation may last from four weeks to the entirety of the unit of work. The outcome of an extended experimental investigation is a written scientific report. *Aspects of each of the three criteria should be evident in the investigation.* For monitoring, the discussion/conclusions/evaluation/recommendations of the report should be between 1500 and 2000 words.

To complete such an investigation the student must:

- develop a planned course of action
- clearly articulate the hypothesis or research question, providing a statement of purpose for the investigation
- provide descriptions of the experiment
- show evidence of modification or student design
- provide evidence of primary and secondary data collection and selection
- execute the experiment(s)
- analyze data
- discuss the outcomes of the experiment
- evaluate and justify conclusion(s)
- present relevant information in a scientific report.

An example from a year 12 paper shows how a student investigated a problem entitled, "The Air Pocket." The assessment starts with a picture of a vertical air jet from a straw producing a cavity on a water surface (see image below).



The student investigated the parameters that would affect the volume of the cavity, preparing a 32-page paper meeting the criteria described earlier, including evaluating the problem theoretically and empirically, presenting data through tables and charts, analyzing findings both by summarizing individual results and developing a regression to evaluate the combined effects of several variables on the volume of the cavity, and evalu-

ating the results, along with the potential errors and additional research needed. Overall, the paper more closely resembles a research report from a scientific laboratory than a traditional high school physics test. The student concluded:

> It was determined through initial theoretical research that the predominant influences on the cavity's volume were air speed, diameter of nozzle/straw and distance between straw/nozzle and water. Upon testing the effects of changing an individual parameter with respect to volume, every possible variation was tried, such that eventually a complete set of values was obtained. To combine the different parameters into a single equation, a multiple regression was used; to determine both the constant factor and the powers to which each of the variables should be raised. The resultant r2 value was 0.96 indicating an excellent fit for the data while the average percentage error was 1.59% and the median percentage error, 6.71%. … [In future experiments], it would be suggested to do the experiments on a larger scale as this would virtually eliminate the effects of surface tension while cutting down unfounded accuracy in the model (the volume could be measured in cubic centimetres or cubic metres, resulting in a more realistic fit, with data that is not required to be impossibly precise. Finally, it would be suggested to trial the effects of the different orientation of the straw/nozzle, as tilting it would give a completely differently shaped cavity (due to the dispersion characteristics of air).

Thus, students go beyond their own empirical data and conclusions to reflect on the accuracy of their findings and means for improving their investigation. These kinds of extended responses are demanded in all of the subject areas, shaped by the core concepts and modes of inquiry of the disciplines. Student reflection is also a common element of the assessments. Consistent scoring of such intellectually ambitious work is made possible in part by internal and external moderation processes (described in the following table), and in part by the clear guidance of the syllabi and rubrics that set standards for the work.

At the end of the year, teachers collect a portfolio of each student's work, which includes the specific assessment tasks, and grade it on a 5-point grading scale. To calibrate these grades, teachers put together a selection of portfolios from each grade level—one from each of the 5 score levels plus borderline cases—and send these to a regional panel for moderation. The panel of five teachers re-scores the portfolios and confers about whether the grade is warranted, making a judgment on the spread. State review panels also look at a sample of student work from each district to insure that schools implement the standards across all districts. Based on this analysis and a 12th grade standardized state-wide test called the Queensland Core Skills (QCS) Test, the Queensland authority confirms the levels of achievement proposed by school programs and may adjust it if it does not calibrate to the standards.

Aiming for even more applied, interdisciplinary work, Queensland developed a "Rich Tasks" approach to standards and assessment, which was introduced as a pilot in 2003.

Physics Objectives

Syllabi are developed to show a clear alignment between the General Objectives (shaded) that state what students should be able to achieve as a result of completing the course and the standards associated with Exit Criteria. Schools must directly teach and assess the General Objectives.

General objectives in the syllabus	Standards				
	The student work has the following characteristics				
Knowledge and conceptual understanding	Standard A	Standard B	Standard C	Standard D	Standard E
Students should acquire knowledge and construct understanding of facts, theories, concepts and principles of physics. To work scientifically, students need to have an understanding of underlying scientific knowledges, including the associated mathematical skills. They need to engage with the processes and phenomena observed in Physics through characteristics of data analysed. Students need to make informed judgments based on sound reasoning in order to direct them in their scientific endeavours and to engage with problem solving. By the end of the course, students should be able to: **recall and interpret concepts, theories and principles of Physics**—this includes the abilities to remember, reproduce and interpret subject matter such as facts, definitions, formulas, terminology, concepts, theories, principles, laws, procedures, sequences, events, diagrams, symbols, figures, systems and patterns **describe and explain processes and phenomena of Physics**—this includes the abilities to compare and classify the concepts, theories and principles being explored, based on primary and secondary data **link and apply algorithms, concepts, theories and schema of Physics**—this includes the abilities to adapt, translate and reconstruct understandings in order to find solutions.	• reproduction and interpretation of complex and challenging concepts, theories and principles • comparison and explanation of complex concepts, processes and phenomena • linking and application of algorithms, concepts, principles, theories and schema to find solutions in complex and challenging situations.	• reproduction and interpretation of complex or challenging concepts, theories and principles • comparison and explanation of concepts, processes and phenomena • linking and application of algorithms, concepts, principles, theories and schema to find solutions in complex or challenging situations.	• reproduction of concepts, theories and principles • explanation of simple processes and phenomena • application of algorithms, principles, theories and schema to find solutions in simple situations.	• reproduction of simple ideas and concepts • description of simple processes and phenomena • application of algorithms, principles, theories and schema.	• reproduction of isolated facts • recognition of isolated simple phenomena • application of simple given algorithms.

(continued on next page)

Physics Objectives (cont'd)

General objectives in the syllabus	Standards				
	The student work has the following characteristics				
Investigative processes	Standard A	Standard B	Standard C	Standard D	Standard E
Students need to recognise the methodologies available to them to investigate scientifically. They need to be able to judge the worth of quantitative and qualitative data and interpret and apply the outcomes of such data. Students require the skills to manipulate and review data and scientific techniques so that they may improve their scientific knowledge. They need to synthesise the research that they have generated and be able to discuss the outcomes in relation to their initial purpose. By the end of the course, students should be able to: **conduct and appraise Physics research tasks**—this includes the abilities to formulate questions, hypothesise, plan, manage, evaluate, refine and justify decisions made during investigations, as well as the critical reflection required to fulfill research goals **operate scientific equipment and technology safely**—this includes the abilities to safely select, adapt and apply technological, laboratory and fieldwork equipment, and consider its limitations; it also incorporates the ability to do this individually and in groups **use primary and secondary data**—this includes the abilities to analyse and extrapolate from data, and to identify relationships, patterns and anomalies in primary and secondary data.	• formulation of justified significant questions/ hypotheses which inform effective and efficient design, refinement and management of investigations • assessment of risk, safe selection and adaptation of equipment, and appropriate application of technology to gather, record and process valid data • systematic analysis of primary and secondary data to identify relationships between patterns, trends, errors and anomalies.	• formulation of justified questions/hypotheses which inform design and management of investigations • assessment of risk, safe selection of equipment, and appropriate application of technology to gather, record and process data • analysis of primary and secondary data to identify patterns, trends, errors and anomalies.	• formulation of questions and hypotheses to select and manage investigations • assessment of risk, safe selection of equipment, and appropriate application of technology to gather and record data • analysis of primary and secondary data to identify obvious patterns, trends, errors and anomalies.	• implementation of given investigations • safe use of equipment and technology to gather and record data • identification of obvious patterns and errors.	• guided use of given procedures • safe directed use of equipment to gather data • recording of data.

(continued on next page)

Physics Objectives (cont'd)

General objectives in the syllabus	Standards				
	The student work has the following characteristics				
Evaluating and concluding	Standard A	Standard B	Standard C	Standard D	Standard E
Students who are working scientifically need to be able to make decisions about the knowledge they have gained and generated. They need to distinguish between a plausible conclusion and one based on pure supposition. Students need to be able to synthesise their thoughts and the thinking of others into a coherent whole, from which they can make judgments and propose future possibilities. They need to reach conclusions and explain the world in which they live, using science. They need to be able to adhere to communication and scientific conventions in communicating their decisions to selected audiences. By the end of the course, students should be able to: **determine, analyse and evaluate the interrelationships involved in applications of Physics**—this includes the abilities to identify the physics involved, to determine the simple and complex relationships that exist between concepts, principles, theories and schema and then to critically examine the associated implications **predict outcomes and justify conclusions and recommendations**—this includes the abilities to explore scenarios and consider possible outcomes, and then to provide justifications of conclusions and recommendations **communicate information in a variety of ways**—this includes the abilities to select, use and present data and ideas to convey meaning, an argument or a case to selected audiences in a range of formats.	The student work has the following characteristics: • analysis and evaluation of complex scientific interrelationships • exploration of scenarios and possible outcomes with justification of conclusions/ recommendations • discriminating selection, use and presentation of scientific data and ideas to make meaning accessible to intended audiences through innovative use of range of formats.	The student work has the following characteristics: • analysis of complex scientific interrelationships • explanation of scenarios and possible outcomes with discussion of conclusions/ recommendations • selection, use and presentation of scientific data and ideas to make meaning accessible to intended audiences in range of formats.	The student work has the following characteristics: • description of scientific interrelationships • description of scenarios and possible outcomes with statements of conclusion/ recommendation • selection, use and presentation of scientific data and ideas to make meaning accessible in range of formats.	The student work has the following characteristics: • identification of simple scientific interrelationships • identification of scenarios or possible outcomes • presentation of scientific data or ideas in range of formats.	The student work has the following characteristics: • identification of obvious scientific interrelationships • statements about outcomes • presentation of scientific data or ideas.

Part of the "New Basics" project, this effort has created extended, multi-disciplinary tasks that are developed centrally and used locally when teachers determine the time is right and they can be integrated with locally-oriented curriculum (Queensland Government, 2001). These are "specific activities that students undertake that have real-world value and use, and through which students are able to display their grasp and use of important ideas and skills." Rich Tasks are defined as:

> A culminating performance or demonstration or product that is purposeful and models a life role. It presents substantive, real problems to solve and engages learners in forms of pragmatic social action that have real value in the world. The problems require identification, analysis and resolution, and require students to analyze, theorize and engage intellectually with the world. As well as having this connectedness to the world beyond the classroom, the tasks are also rich in their application: they represent an educational outcome of demonstrable and substantial intellectual and educational value. And, to be truly rich, a task must be transdisciplinary. Transdisciplinary learnings draw upon practices and skills across disciplines while retaining the integrity of each individual discipline.

One task description is shown on page 17.

A bank of these tasks now exists across grade levels, along with scoring rubrics and moderation processes by which the quality of the tasks, the student work, and the scoring can be evaluated. Extensively researched, this system has had excellent success as a tool for school improvement. Studies found students are more engaged in learning in schools using the Rich Tasks. On traditional tests, these "New Basics" students scored about the same as students in the traditional program, but they performed notably better on assessments designed to gauge higher-order thinking.

The Singapore government has employed the developers of the Queensland system to expand and refine its system of performance assessments. High-scoring Hong Kong has also begun to expand its ambitious school-based assessment system in collaboration with Queensland assessment developers.

Victoria

In Victoria, a mixed system of centralized and decentralized assessment combines these kinds of school-based assessment practices with a set of state exams guided by the Victoria Essential Learning Standards. The AIM program, used at years 3, 5, 7, and 9, indicates how well the literacy and numeracy skills of students are developing. The results provide information used to plan new programs and useful feedback to students, parents, and teachers. Assessment tasks include extended open-ended writing responses and some multiple-choice responses.

At the secondary level, the Victorian Certificate of Education (VCE) provides information that guides pathways to further study at the university, Technical and Further

Science and Ethics Confer

Students must identify, explore and make judgments on a biotechnological process to which there are ethical dimensions. Students identify scientific techniques used as well as significant recent contributions to the field. They will also research frameworks of ethical principles for coming to terms with an identified ethical issue or question. Using this information they prepare pre-conference materials for an international conference that will feature selected speakers who are leading lights in their respective fields.

In order to do this students must choose and explore an area of biotechnology where there are ethical issues under consideration and undertake laboratory activities that help them understand some of the laboratory practices. This enables them to:

A) Provide a written explanation of the fundamental technological differences in some of the techniques used, or of potential use, in this area (included in the pre-conference package for delegates who are not necessarily experts in this area).

B) Consider the range of ethical issues raised in regard to this area's purposes and actions, and scientific techniques and principles and present a deep analysis of an ethical issue about which there is a debate in terms of an ethical framework.

C) Select six real-life people who have made relevant contributions to this area and write a 150-200 word précis about each one indicating his/her contribution, as well as a letter of invitation to one of them.

This assessment measures research and analytic skills; laboratory practices; understanding biological and chemical structures and systems, nomenclature and notations; organizing, arranging, sifting through, and making sense of ideas; communicating using formal correspondence; précis writing with a purpose; understanding ethical issues and principles; time management, and much more.

Education (TAFE) and to the world of work. Some students undertake a school-based apprenticeship or traineeship within the VCE. The Victoria Curriculum and Assessment Authority establishes courses in a wide range of studies, develops the external examinations, and ensures the quality of the school-assessed component of the VCE.

VCAA conceptualizes assessment as "of," "for," and "as" learning. Teachers are involved in developing assessments, along with university faculty in the subject area. All prior year assessments are public, in an attempt to make the standards and means of mea-

suring them as transparent as possible. Before the external examinations are given to students, teachers and academics sit and take the exams themselves, as if they were students. The external subject-specific examinations, given in grades 11 and 12, include about 25% machine-scored items; the remaining items are open-ended and are scored by the classroom teacher. The exams may include written, oral, and performance elements.

In addition, classroom-based tasks that are given throughout the school year comprise at least half of the total examination score. Teachers design these required assignments and assessments—lab experiments and investigations on central topics as well as research papers and presentations—in response to syllabus expectations. The required classroom tasks ensure that students are getting the kind of learning opportunities which prepare them for the assessments they will later take, that they are getting feedback they need to improve, and that they will be prepared to succeed not only on these very challenging tests but in college and in life.

An example from the Victoria biology test (page 19), for example, describes a particular virus to students, asks them to design a drug to kill the virus and, in several pages, explain how the drug operates (complete with diagrams), and then to design an experiment to test the drug.

In preparation for this on-demand test, students taking biology will have been assessed on six pieces of work during the school year covering specific outcomes in the syllabus. For example, they will have conducted "practical tasks" such as using a microscope to study plant and animal cells by preparing slides of cells, staining them, and comparing them in a variety of ways, resulting in a written product with visual elements. They also will have conducted practical tasks on enzymes and membranes, and on the maintenance of stable internal environments for animals and plants. Finally, they will have completed and presented a research report on characteristics of pathogenic organisms and mechanisms by which organisms can defend against disease. These tasks link directly to the expectations that students will encounter on the external examination but go well beyond what that examination can measure in terms of how students can apply their knowledge.

The tasks are graded according to criteria set out in the syllabus. The quality of the tasks assigned by teachers, the work done by students, and the appropriateness of the grades and feedback given to students are audited through an inspection system, and schools are given feedback on all of these elements. In addition, the VCAA uses statistical moderation to ensure that the same assessment standards are applied to students across schools. The external exams are used as the basis for this moderation, which adjusts the level and spread of each school's assessments of its students to match the level and spread of the same students' collective scores on the common external test score. The result is a rich curriculum for students with extensive teacher participation and a comparable means for examining student learning.

Victoria, Australia, High School Biology Exam

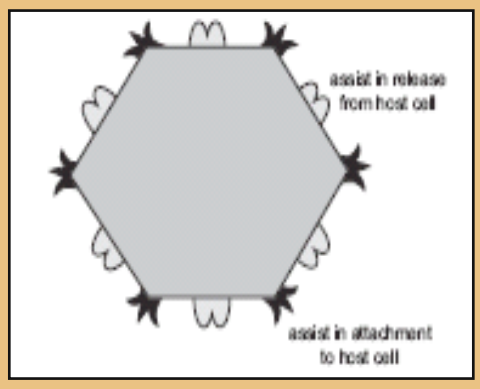
When scientists design drugs against infectious agents, the term "designed drug" is often used.

A. Explain what is meant by this term:

_____

_____

Scientists aim to develop a drug against a particular virus that infects humans. The virus has a protein coat and different parts of the coat play different roles in the infective cycle. Some sites assist in the attachment of the virus to a host cell; others are important in the release from a host cell. The structure is represented in the following diagram:



The virus reproduces by attaching itself to the surface of a host cell and injecting its DNA into the host cell. The viral DNA then uses the components of host cell to reproduce its parts and hundreds of new viruses bud off from the host cell. Ultimately the host cell dies.

B. Design a drug that will be effective against this virus. In your answer outline the important aspects you would need to consider. Outline how your drug would prevent continuation of the cycle of reproduction of the virus particle. Use diagrams in your answer. Space for diagrams is provided on the next page.

C. Before a drug is used on humans, it is usually tested on animals. In this case, the virus under investigation also infects mice. Design an experiment, using mice, to test the effectiveness of the drug you have designed.

Hong Kong

In collaboration with educators from Australia, the United Kingdom, and other nations, Hong Kong's assessment system is evolving from a highly centralized examination system to one that increasingly emphasizes school-based, formative assessments that expect students to analyze issues and solve problems. While school-based assessments have been part of Hong Kong's examination system since 1978, they are assuming greater prominence in the government's plan to gradually replace the Hong Kong

Certificate of Education Examinations, which most students sit for at the end of their five-year secondary education, with a new Hong Kong Diploma of Secondary Education that will combine on-demand tests with curriculum-embedded tasks. . In addition, the Hong Kong Territory-wide System Assessment (TSA), which assesses lower-grade student performance in Chinese, English, and mathematics, is developing an online bank of assessment tasks to enable schools to assess their students and receive feedback on their performance on their own timeframes. The formal TSA assessments, which include both written and oral components, occur at Primary Grades 3 and 6 and Secondary Grade 3 (the equivalent of grade 9 in the United States).

As outlined in Hong Kong's "Learning to learn" reform plan, the goal of the reforms is to shape curriculum and instruction around critical thinking, problem-solving, self-management skills, and collaboration. Another key goal is to help students develop meta-cognitive thinking skills, so they can identify their strengths and areas needing additional work (Education Bureau, September 2001; Chan, et al., 2008).

The Hong Kong Education Examinations Authority explains the rationale for growing use of school-based assessments (SBA):

> The primary rationale for SBA is to enhance the validity of the assessment, by including the assessment of outcomes that cannot be readily assessed within the context of a one-off public examination. SBA can also reduce dependence on the result of public examinations, which may not always provide the most reliable indication of the actual abilities of candidates. Obtaining assessments based on student performance over an extended period of time and developed by those who know the students best - their subject teachers - provides a more reliable assessment of each student. Another reason for including SBA is to promote a positive impact on teaching and learning. It can serve to motive students by engaging them in meaningful activities; and for teachers, it can reinforce curriculum aims and good teaching practices, and provide structure and significance to an activity that they are in any case involved in on a daily basis, namely assessing their own students….
>
> Teachers know that SBA, which typically involves students in activities such as making oral presentations, developing a portfolio of work, undertaking fieldwork, carrying out an investigation, doing practical laboratory work or completing a design project, help students to acquire important skills , knowledge and work habits that cannot readily be assessed or promoted through paper-and-pencil testing. Not only are they outcomes that are essential to learning within the disciplines, they are also outcomes that are valued by tertiary institutions and by employers. Moreover, they are activities that students find meaningful and enjoyable. (HKEAA, 2009).

By 2007, Curriculum and Assessment Guides were published for four core subjects and 20 elective subjects, and assessments in the first two subjects—Chinese language and English language—were revised. The language tests became criterion-referenced, performance-based assessments featuring not only essays previously used on the exams, but also new speaking and listening components, the composition of written papers testing integrated skills, and a school-based component generally worth 20 to 30% of the examination score. Although the existing assessments already use open-ended responses (see the example of a physics examination question in Appendix B), the proportion of such responses will increase in the revised assessments.

These new assessments feature "standards-referenced reporting," in which examinations are scored on a scale from 1 to 5 pegged to specific criteria, rather than the norm-referenced approach that produced earlier grades. The Hong Kong Examinations and Assessment Authority explains that, "By providing both students and teachers with a specific set of level descriptors, SRR not only clarifies the standards expected at various levels of attainment but also helps set targets for improvements in learning and teaching" (HKEAA, 2007).

Like the existing assessments, teachers develop the new assessments with the participation of higher education faculty, and they are scored by teachers who are trained as assessors. Tests are allocated randomly to scorers, and essay responses are typically rated by two independent scorers (Dowling, n.d.). Results of the new school-based assessments are statistically moderated to ensure comparability within the province. The assessments are internationally benchmarked, through the evaluation of sample student papers, to peg the results to those in other countries. Many of the new assessments will also be scored online, a practice which the Examinations Authority notes is now common in 20 of China's mainland provinces, as well as in the United Kingdom.

The Education Bureau encourages schools to develop assessments focused on learning, and is shifting its education policies to underscore this focus. For example, the Bureau promotes the use of multiple forms of assessment in schools including projects, portfolios, observations, and examinations, and looks for the variety of assessments in the performance indicators used for school evaluation (Chan, et al., 2008; Quality Assurance Division of the Education Bureau, 2008).

To guide the process of assessment reform, the Education Bureau implemented a School Development and Accountability Framework in the 2003-2004 school year. The framework emphasizes school self-evaluation, as well as an external peer evaluation, using a set of performance indicators. For example, with respect to curriculum and assessment, the performance indicators ask: "Is the school able to adopt varied modes of assessment and effectively assess students' performance in respect of knowledge, skills, and attitude?" and "How does the school make use of curriculum evaluation data to inform curriculum planning?" (Quality Assurance Division of the Education Bureau, 2008). The Education Bureau also conducts a Quality Assurance Inspection, in which personnel from other

schools and from the Education Bureau conduct on-site examinations of how each school meets the performance indicators.

Singapore

Singapore is also emphasizing the integration of school-based assessment into large-scale testing systems. Policy analysts have been intensely interested in Singapore's education system since its students took first place in the Trends in International Mathematics and Science Study (TIMSS) assessments in mathematics and science in 1995, 1999, and 2003. Singapore's scores are based on high achievement by all of the country's students, including the Malay and Tamil minorities, who have been rapidly closing what was once a yawning achievement gap (Dixon, 2005). About 90% of Singapore's students scored above the international median on the TIMSS tests. This accomplishment is even more remarkable given that fewer than half of Singapore's students routinely speak English, the language of the test, at home. Most speak one of the other four official national languages of the country—Mandarin, Malay, or Tamil—and some speak one of dozens of other languages or dialects.

Intensive investment and reform over a period of 30 years have transformed the Singaporean education system, broadening access and increasing equality while orchestrating a system that includes a complex system of private, "autonomous," and public schools. Some of the schools are inherited from the colonial era, and all receive government subsidies. These schools are intentionally diverse in many ways. Local schools are urged to innovate, but they are similar in instructional expectations and supports, and use a common national curriculum for core subjects.

Since the prime minister introduced the "thinking schools, learning nation" initiative in 1997, Singapore's explicit focus in its reforms of curriculum, assessment, and teaching has been to develop a creative and critical thinking culture within schools. The goal is to teach and assess these skills for students and to create an inquiry culture among teachers as well. Teachers are encouraged to conduct action research on their teaching and to continually revise their teaching strategies in response to what they learn. This initiative was married to commitments to integrate technology in all aspects of education—a mission nearly fully accomplished a decade later—and to dramatically open up college and university admissions.

Higher education is now available to virtually all Singaporeans. Based on their interests, labor force needs, and the results of their grades, O-level exams, and other accomplishments, students pursue one of three pathways after 10th grade, when secondary school ends. About 25% attend junior college for two years, followed by university, which leads to professional paths such as teaching, science, engineering, medicine, law, and the civil service; about 60% attend a polytechnic college for three years, after which about half go on to university while the others go into jobs in technical and engineering fields; the remainder—about 15%—attend an Institute of Technical Education for two years,

after which some continue on to college or university. Virtually everyone finishes one of these pathways.

Historically, Singapore has operated a modified British-style system of examinations. Students sit for national exams administered by the Singapore Examinations and Assessment Board (SEAB). At the end of Year 6 (age 12), students take the Primary School Leaving Examinations (PSLE). These are open-ended written and oral examinations in four core subject areas—mathematics, science, English and a "mother tongue" language—that are administered and scored by teachers in moderated scoring sessions. The exams in the English and native languages include four components—two written essays of at least 150 words, listening comprehension, language comprehension, and an oral exam that requires students to engage in a conversation on a set topic for 15 minutes. Two examiners observe the candidates and grade the oral proficiency of the student. In math, students have to demonstrate the steps in solving a problem.

Students take the General Certificate of Examinations Normal or Ordinary Level (GCE N/O-Level) at the end of year 10 (age 16). The GCE O-level examinations are based on common course syllabi that outline what is to be taught; they require short and long open-ended responses and essays across a wide range of content areas. Students choose the areas in which they want to be examined. Although the results are used to guide postsecondary admissions, and are not used to determine graduation from high school, they exert substantial influence on the high school curriculum. Recent reforms are changing the curriculum and assessment system to make it more explicitly focused on creativity and independent problem solving.

Students attending Junior College (grades 11 and 12) en route to university take the GCE Advanced Level (A-Level) exams at the end of year 12 (age 18). A new 'A' level curriculum and examination system was introduced in 2002. The new exams are meant to encourage multi-disciplinary learning by requiring that students "select and draw together knowledge and skills they have learned from across different subject areas, and apply them to tackle new and unfamiliar areas or problems" (Singapore Examinations and Assessment Board, 2006, p. 2).

The A-level curricular framework includes Core Content Areas in which students take courses and associated exams: humanities, mathematics and sciences, and languages. It also includes Life Skills —emphasizing leadership, enrichment, and service to others— and Knowledge Skills, evaluated through a general paper, project work, and a course in knowledge and inquiry. A typical A-level student is evaluated in three compulsory subjects—a general paper, project work, and a native language assessment—along with four content subjects.

The newer areas of Life Skills and Knowledge Skills are intended to develop the more advanced thinking skills thought to be underrepresented in the traditional content-based curriculum and examinations system. They represent the goals of reforms

launched in 1997 as part of the "thinking schools, learning nation" initiative, which created a number of changes:

> Syllabi, examinations and university admission criteria were changed to encourage thinking out of the box and risk-taking. Students are now more engaged in project work and higher order thinking questions to encourage creativity, independent, and inter-dependent learning (Ng, 2008, p. 6).

The content courses are also evolving to include more critical thinking, inquiry, and investigation, along with mastery of content. A number of the high school content tests are accompanied by school-based tasks, such as research projects and experiments designed and conducted by students. Each of the science courses now includes a component called the "School-based Science Practical Assessment." These school-based components, which teachers manage and score according to specifications provided by the Examinations Board, count for up to 20% of the examination grade. Scoring is both internally and externally moderated. The goal is for students to be able to:

1. Follow a detailed set or sequence of instructions and use techniques, apparatus, and materials safely and effectively;
2. Make and record observations, measurements, methods, and techniques with precision and accuracy;
3. Interpret and evaluate observations and experimental data; and
4. Identify a problem, design and plan investigations, evaluate methods and techniques, and suggest possible improvements in the design.

The projects can be submitted to the university as part of the application, and universities are encouraged to examine evidence about student accomplishments beyond examination scores. On pages 25-26 we describe some of these innovations in the examination system.

Intellectually challenging school-based assessments are also encouraged in the earlier grades. The curriculum and assessment guidelines that accompany the national standards suggest that teachers engage in continual assessment in the classroom using a variety of assessment modes, such as classroom observations, oral communication, written assignments and tests, and practical and investigative tasks. The Ministry has developed a number of curriculum and assessment supports for teachers. For example, SAIL (Strategies for Active and Independent Learning) aims to support more learner-centered project work in classrooms and provides assessment rubrics to clarify learning expectations. All schools have received training for using these tools.

The Ministry's 2004 Assessment Guides for both primary and lower secondary mathematics contain resources, tools, and ideas to help teachers incorporate strategies such as mathematical investigations, journal writing, classroom observation, self-assessment,

Project Work

Project Work (PW) is **an interdisciplinary subject** that is compulsory for all pre-university students. There is dedicated curriculum time for students to carry out their project tasks over an extended period. As an interdisciplinary subject, it breaks away from the compartmentalization of knowledge and skills to focus on interdisciplinary outcomes by requiring students to draw knowledge and apply skills from across different subject domains. The goals for this experience are embedded in the requirements for the task and its assessment, which are centrally set by the Singapore Examinations and Assessment Board. The tasks are designed to be sufficiently broad to allow students to carry out a project that they are interested in while meeting the task requirements:

It must foster collaborative learning through group work. Together as a group, which is randomly formed by the teacher, students brainstorm and evaluate each others' ideas, agree on the project that the group will undertake, and decide on how the work should be allocated amongst themselves.

Every student must make an oral presentation: Individually and together as a group, each student makes an oral presentation of his/her group project in the presence of an audience.

Both product and process are assessed: There are three components for assessment:

- the **Written Report** which shows evidence of the group's ability to generate, analyze and evaluate ideas for the project;
- the **Oral Presentation** in which each individual group member is assessed on his/her fluency and clarity of speech, awareness of audience as well as response to questions. The group as a whole is also assessed in terms of the effectiveness of the overall presentation;
- the **Group Project File** in which each individual group member submits three documents related to 'snapshots' of the processes involved in carrying out the project. These documents show the individual student's ability to generate, analyze, and evaluate (i) preliminary ideas for a project, (ii) a piece of research material gathered for the chosen project, and (iii) insights and reflections on the project.

In carrying out the PW assessment task, students are intended to acquire self-directed inquiry skills as they propose their own topic, plan their timelines, allocate individual areas of work, interact with teammates of different abilities and personalities, gather and evaluate primary and secondary research material. These PW processes reflect life skills and competencies such as knowledge application, collaboration, communication and independent learning, which prepare students for the future workplace.

About 12,000 students complete this task annually. Assessment is school-based and criterion-referenced. While the SEAB externally specifies task setting, conditions, assessment criteria, achievement standards, and marking processes, classroom teachers carry out the assessment of all three components of PW using a set of assessment criteria provided by the board. All schools are given exemplar material that illustrates the expected marking standards. The Board provides training for assessors and internal moderators. Like all other assessments, the grading is both internally and externally moderated.

Knowledge and Inquiry

Knowledge and Inquiry is a Humanities subject that seeks to develop in students

An understanding of the nature and construction of knowledge: Students are expected to show that they have read widely and have understood and can apply the concepts they encountered. They are expected to demonstrate skill in selecting relevant material with which to tackle the assessment tasks.

Critical Thinking: Students are expected to demonstrate that they can analyze different kinds of arguments and information, identify and evaluate assumptions and points of view, verify claims, and provide reasoned and supported arguments of their own.

Communication Skills: Students are expected to communicate their ideas and arguments clearly and coherently in good English. They are expected to structure their arguments, select an appropriate style of presentation that is fully relevant to the questions asked, and demonstrate clear ability to engage with different aspects of the questions.

There are three assessment components:

Essay: This paper gives candidates the opportunity to demonstrate their ability to apply the concepts they have learned in their study of the nature and construction of knowledge.

It covers the theoretical aspects of areas of exploration identified in the syllabus and the question set requires candidates to draw on knowledge they have gained during their study of the following key questions:

Why ask questions?
What is knowledge?
How is knowledge constructed?
What makes knowledge valid?
How is knowledge affected by society?
How should knowledge be used?

Critical Thinking: This paper requires students to critically analyze different kinds of arguments and information presented in the material, identify and evaluate assumptions and points of view, verify claims, and provide reasoned and supported arguments. Students must use language appropriately and effectively to communicate a clear and well-structured argument.

Independent Study: The independent study component allows students to demonstrate their understanding of the nature and construction of knowledge as it relates to their chosen area of study, apply this understanding in addressing the specific context, select appropriate material and show that they have engaged in relevant reading during the course of their research by presenting a literature review, and apply what they have read to support the arguments they present. Students must use language appropriately and effectively to communicate a clear and well-structured argument. At the end of the six months of independent research study, they submit an extended essay of 2,500 to 3,000 words.

and portfolio assessment into the classroom. Emphasis is placed on the assessment of problem solving and on meta-cognition, the self-regulation of learning that enables students to internalize standards and become independent learners (Kaur, 2005). The Institute of Education has held workshops to support learning about the new assessments and integrated the new strategies into teacher development programs.

United Kingdom

The move toward more school-based assessment has also occurred in various ways in the United Kingdom, which for more than a century has influenced examination systems in English-speaking countries around the world. Assessments have typically been open-ended essay and constructed-response examinations, but the nature of the tasks and of the administration has changed over the last two decades to include more school-based tasks and projects.

England

England's assessment system is managed at the national level by an organization called the Qualifications and Curriculum Authority (QCA). Schools teach and assess students using a national curriculum, which includes syllabi for specific courses.

Teachers assess pupils' progress continuously and assemble evidence for external reporting in the national data system at ages 7, 11, and 14 (Key Stages 1, 2, and 3). This evidence is based on classroom-based assignments, observations, and tasks, the results of which are evaluated in terms of indicators of performance outlined in learning progressions for each of several dimensions of learning within each subject area.

At Key Stage 1, student progress is evaluated based on classroom evidence and results from centrally-developed, open-ended tests and tasks in English and mathematics. The tests and tasks are marked by teachers and moderated within the school and by external moderators. At Key Stage 2, student progress is evaluated based on teachers' summary judgments and results from open-ended tests in English, mathematics, and science. These tests are externally marked and the results reported on a national level. For Key Stage 3, England recently abolished external tests and now relies on teacher assessments to report achievement levels in all subjects. Teacher judgments are moderated and results are reported on a national level.

The Assessing Pupils' Progress program that guides this work is described by the QCA in this way:

> APP is the new structured approach to teacher assessment, developed by QCA in partnership with the National Strategies, which equips teachers to make judgments on pupils' progress. It helps teachers to fine-tune their understanding of learners' needs and to tailor their planning and

> teaching accordingly, by enabling them to: use diagnostic information about pupils' strengths and weaknesses to improve teaching, learning and pupils' progress; make reliable judgments related to national standards drawing on a wide range of evidence; and track pupils' progress.
>
> The APP subject materials for teachers include assessment guidelines for assessing pupils' work in relation to national curriculum levels. These provide a simple recording format providing assessment criteria for each of the assessment focuses in the subject, and standards files, which are annotated collections of pupils' day-to-day work that exemplify national standards at different levels. These help teachers reach consistent and reliable judgments about national curriculum levels (Qualifications and Curriculum Authority, 2009, p. 1.)

Some nationally-developed tasks are designed and distributed to schools to support teacher assessment. At Key Stage 2 (age 11), a set of these tasks and tests must be used to evaluate students in combination with the other evidence teachers assemble from the classroom. In other years, the use of the tasks is optional. As described by the QCA:

> The tasks are designed to support teacher assessment. They can be used to indicate what pupils are able to do and inform future learning and teaching strategies. Individual tasks can be used to provide a basis for discussion by teachers and pupils on what has been achieved and to identify the next steps. They can support day-to-day assessment and generate outcomes which can contribute to the breadth of evidence which is used as the basis for periodic and transitional assessment.

At Key Stage 4, ages 15 to 16, the national qualification framework includes multiple pathways for students and consequently multiple measures of student achievement. There are four pathways based on students' aspirations after graduation: apprenticeship, diploma, the General Certificate of Secondary Education (GCSE), and the A-level examinations. Some students go on to a Further Education college to take vocationally related courses. They usually take the National Vocational Qualification using the apprenticeship model.

Most students take the GCSE, a two-year course of study evaluated by assessments both within and at the end of courses or unit. Students may take as many single-subject or combined-subject assessments as they like, and they choose which ones they will take based on their interests and areas of expertise. The exams involve constructed response items and structured, extended classroom-based tasks which comprise from 25 to 60% of the final examination score. England is currently piloting new tasks for the GCSE with an increased emphasis on functional skills like problem solving, team building, and communication as well as personal learning and thinking skills across subjects. These new tasks, called "controlled assessments" are either designed by the award-

ing body and marked by teachers or designed by teachers and marked by the awarding body. Either way teachers determine the timing of controlled assessments.

These classroom-based assessments comprise 25% of the total examination score in subjects such as business studies, classical civilization, English literature, geography, history, humanities, or statistics, and 60% of the total examination score in subject areas such as applied business, music and dance, design and technology, drama, engineering, English Language, expressive arts, health and social care, home economics, Interactive Computer Technology (ICT), manufacturing, media studies, and modern foreign languages. Following are examples of classroom-based tasks in English and ICT.

Example of Tasks: GCSE English	
Unit and Assessment	Tasks
Reading literacy texts Controlled assessment (coursework) 40 marks	Responses to three texts from choice of tasks and texts. Candidates must show an understanding of texts in their social, cultural and historical context.
Imaginative Writing Controlled assessment (coursework) 40 marks	Two linked continuous writing responses from a choice of Text Development or Media.
Speaking and Listening Controlled assessment (coursework) 40 marks	Three activities: a drama-focused activity; a group activity; an individual extended contribution. One activity must be a real-life context in and beyond the classroom.
Information and Ideas Written exam 80 marks (40 per section)	Non-Fiction and Media: Responses to unseen authentic passages. Writing information and Ideas: One continuous writing response—choice from two options.

During Key Stage 4, most students take five or more GCSE exams. Their performance determines the level of the diploma they receive, and whether they will go on to Advanced Studies, which are evaluated by A-level exams that qualify students for university admissions. England has 45 areas for A-level exams. The exam questions require extended answers aimed at assessing deeper levels of understanding and applications of knowledge to real-world problems, as illustrated in the example on page 31.

Most of the exams take the form of essay questions. The mathematics exams include questions that ask students to show the reasoning behind their answers. Foreign language exams require oral presentations. The A-level exam in English literature asks students to show their skills and knowledge in four sections: poetry, drama, prose, and general. Students analyze works of literature they have read as part of their curriculum in terms of their meaning and interpretation as well as literary devices and writing strategies. Coursework accounts for 25 to 30% of the A-level score, depending on the course.

GCSE Controlled Assessment Task in Interactive Computer Technology

Litchfield Promotions works with more than 40 bands and artists to promote their music and put on performances in England. The number of bands they have on their books is gradually expanding. Litchfield Promotions needs to be sure that each performance will make enough money to cover all the staffing costs and overheads as well as make a profit. Many people need to be paid: the bands; sound engineers; and lighting technicians. There is also the cost of hiring the venue. Litchfield Promotions needs to create an Interactive Computer Technology solution to ensure that they have all necessary information and that it is kept up to date. Their solution will show income, outgoings and profit.

Candidates will need to: 1) Work with others to plan and carry out research to investigate how similar companies have produced a solution. The company does not necessarily have to work with bands and artists or be a promotions company. 2) Clearly record and display your findings. 3) Recommend a solution that will address the requirements of the task. 4) Produce a design brief, incorporating timescales, purpose and target audience.

Produce a solution, ensuring that the following are addressed: 1) It can be modified to be used in a variety of situations. 2) It has a friendly user interface. 3) It is suitable for the target audience. 4) It has been fully tested. You will need to: 1) incorporate a range of: software features, macros, modeling, and validation checks - used appropriately. 2) Obtain user feedback. 3) Identify areas that require improvement, recommending improvement, with justification. 4) Present information as an integrated document. 5) Evaluate your own and others' work.

Students must now also complete an independently-designed extended research project as part of the A-level assessments. Assessments are marked by teachers in a moderated process managed by the five examination agencies that organize sets of examinations.

While England has moved to include some school-based assessments in its increasingly performance-oriented assessment system, Scotland, Wales, and Northern Ireland have gone even further in revising their approaches to assessment.

Scotland

Scotland has a separate governing body for its educational system from the United Kingdom. An assessment called the Scottish Survey of Achievement is administered in the third, fifth, and seventh years of primary school. Standardized courses and benchmark exams are used in secondary school. Teachers and lecturers design and mark the assessment tasks for the primary courses and general secondary courses, and schools use external assessments for the intermediate and advanced secondary courses. The

Sample A-Level Question from a Probability and Statistics 1 Exam

A city council attempted to reduce traffic congestion by introducing a congestion charge. The charge was set for 4 pounds for the first year and was then increased by 2 pounds each year. For each of the first eight years, the council recorded the average number of vehicles entering the city center per day. The results are shown in the table:

Charge, (pounds) *x*	4	6	8	10	12	14	16	18
Average number of vehicles per day, *y* million	2.4	2.5	2.2	2.3	2.0	1.8	1.7	1.5

1. Calculate the product moment correlation coefficient for these dates.
2. Explain why x is the independent variable.
3. Calculate the equation of the regression line of *y* on *x*.

4a. Use your equation to estimate the average number of vehicles, which will enter the city center per day when the congestion charge is raised to 20 pounds.

4b. Comment on the reliability of your estimate.

5. The council wishes to estimate the congestion charge required to reduce the average number of vehicles entering the city per day to 1.0 million. Assuming that a reliable estimate can be made by extrapolation, state whether they should use the regression line of *y* on *x* or the regression line of *x* on *y*. Give a reason for your answer.

Scottish Qualifications Authority designs and scores those assessments, which may take the form of an examinations, project work, or portfolios (Scottish Qualifications Authority, March 2004; The Scottish Government, 2008).

Wales

Wales only recently separated from the system used in the United Kingdom and now has a separate governing body for its educational system (Archer, 2006). The more centralized system introduced in England under the Thatcher administration (later modified during the Blair administration as described above) inspired policies like No Child Left Behind in the United States, but caused consternation among countries in the United Kingdom that favored a different approach. Wales broke from the British system and opted to abolish national exams for children through age 14.

Much like Finland, during the primary years Welsh schools have a national school curriculum supported by teacher-created, administered, and scored assessments. During

the secondary years, teachers create and manage all assessment of 14-year-old students, while students 16 years and older are encouraged to participate in the relevant GCSE exams and A-level courses and exams administered by the England's Qualifications and Curriculum Authority (Welsh Assembly Government, 2008a and 2008b). With these changes to its assessment system, Wales hopes to increase student engagement and reduce teaching to the test (Archer, 2006).

Northern Ireland

Northern Ireland is in the process of implementing an approach at all levels called "Assessment for Learning." This approach emphasizes locally developed, administered and scored assessments and focuses on five key actions:

1. *Sharing learning intentions* where students and teacher agree upon learning intentions to give them ownership over their learning.
2. *Sharing and negotiating success criteria* where students and teacher create the criteria for successful completion of a task together to help with self-assessment.
3. *Feedback* where teachers provide ongoing feedback during formative assessment sessions.
4. *Effective questioning* where teachers introduce strategies like using open-ended questions and giving more thinking time so students will feel more confident thinking aloud and explaining their reasoning.
5. *How pupils reflect on their learning* where teachers provide students with strategies to think about what they have learned.

Northern Ireland does not require schools to externally assess students up through age 14, but it provides teachers with the option to give students assessments at the end of Stage 3, which are externally graded through the Northern Ireland Council for the Curriculum Examinations and Assessments (CCEA). The largely open-ended assessments evaluate how students reason, think, and problem solve. CCEA provides multiple assessments for Stage 4, according to which pathway a student chooses to follow, including taking the GCSE exam and A-level courses and exams from the U.K. system (i.e. whether towards university or a vocational degree) (Council for the Curriculum Examinations and Assessment, 2008a and b).

International Baccalaureate Diploma Program

The International Baccalaureate Diploma Program (IB) is offered in 1,600 schools serving 100,000 students in 125 countries. Like systems in many high-achieving nations, the IB high school curriculum is standards-based and syllabus-based, integrating assessment within the curriculum in a tightly constructed teaching and learning system that blends classroom-based and external examinations. The IB program generally serves students in grades 11 and 12, assessing students using school-based assessments during the two-year program and externally-based exams at the end of the course of study.

Both types of assessments measure students' individual performance on the objectives outlined by syllabi, or "subject outlines" written by the International Baccalaureate Organization (IBO), the non-profit educational foundation that provides services and support to IB schools worldwide.

In almost all subjects, teachers conduct school-based assessments by grading individual pieces of coursework based on the objective set out by the IB subject outlines. School-based assessments contribute between 20 and 30% of the total grade in most subjects and as much as 50% in arts courses like music, theater arts, and visual arts. Coursework graded by teachers includes such assessments as oral exercises in language subjects, projects, student portfolios, class presentations, practical laboratory work, mathematical investigations, and artistic performances (International Baccalaureate Organization, 2008).

The externally-based exams usually consist of essays, structured problems, short-response questions, data-response questions, text-response questions, case-study questions, and a limited use of multiple choice questions. There are a limited number of externally assessed pieces of work (i.e., a theory of knowledge essay, extended essay, and world literature assignment) that students complete over an extended period of time under teacher supervision, but which are marked by external evaluators, or "IB Examiners," personnel trained and organized by the IBO.

IB externally-based exams ask students to apply the analytical and problem solving skills and content knowledge they gain in their IB coursework to specific problems or analyses. For example, a sample "English A1—Higher Level" essay asks students to answer one essay question and base their answer on a least two of three works studied in class. Students pick from 5 categories: drama, poetry, prose: the novel and short story, prose: other than the novel and short story, or general questions on literature.

On the English exam, students may choose from essay questions like the following:

1. Using two or three of the works you have studied, discuss how and to what effect writers have used exaggeration as a literary device.
2. Acquiring material wealth or rejecting its attractions has often been the base upon which writers have developed interesting plots. Compare the ways the writers of two or three works you have studied have developed such motivations.
3. Discuss and compare the role of the speaker or persona in poems you have studied. You must refer closely to the work of two or three poets in your study and base your answer on a total of three or four poems (International Baccalaureate Organization, 2005).

A typical "Mathematics Standard Level" essay asks students to show their work and support their answers with work and explanations. It also asks students to draw any

graphs they create on their graphing calculator, noting that students may receive credit for using the correct method even if an error produces an incorrect final answer. The exam includes five multi-step questions. The following question addresses multiple math skills including probability, proportions, and algebra:

> In a large school, the heights of all 14-year-old students are measured. The heights of the girls are normally distributed with mean 155cm and standard deviation 10cm. The heights of the boys are normally distributed with mean 160cm and standard deviation 12cm.
>
> a) Find the probability that a girl is taller than 170 cm.
> b) Given that 10% of the girls are shorter than x cm, find x.
> c) Given that 90% of the boys have heights between q cm and r cm where q and r are symmetrical about 160 cm, and $q < r$, find the value of q and of r.
>
> In the group of 14-year-old students, 60% are girls and 40% are boys. The probability that a girl is taller than 170 cm was found in part (a). The probability that a boy is taller than 170 cm is 0.202. A 14-year-old student is selected at random.
>
> d) Calculate the probability that the student is taller than 170 cm.
> e) Given that the student is taller than 170 cm, what is the probability the student is a girl?
>
> (International Baccalaureate Organization, 2006).

In addition to these kinds of assessments of specific course-based learning, the IB course of studies, like the British A-levels and the Singaporean exams, requires students to complete an extended essay. This paper is an independent, self-directed piece of research, culminating in a 4,000-word paper. According to the IB organization, it is intended to provide practical preparation for the kinds of undergraduate research required in college, and is an opportunity for students to engage in an in-depth study of a topic of interest within a chosen subject. Emphasis is placed on the research process: formulating an appropriate research question, engaging in a personal exploration of the topic, communicating ideas, and developing an argument. Participation in this process is intended to develop students' capacity to analyze, synthesize, and evaluate knowledge.

Conclusion

An examination of assessment practices in a number of high-achieving nations and rigorous programs such as the International Baccalaureate illustrates the importance of assessment *of, for* and *as* learning, rather than as a separate disjointed element of the education enterprise. High-quality assessments provide feedback to students, teachers, and schools about what has been learned, and they "feed-forward" information that can shape future learning, as well as guide college and career-related decision making.

These systems closely align curriculum expectations, subject and performance criteria, and desired learning outcomes. They engage teachers in assessment development and scoring as a way to improve their professional practice and their capacity to support student learning and achievement. They engage students in authentic assessments to improve their motivation and learning. They seek to advance student learning in higher-order thinking skills and problem-solving by using a wider range of instructional and assessment strategies. And they privilege quality over quantity of standardized testing—moving systems from 'accounting' to more useful 'accountability' for learning.

References

Archer, J. (2006, December 19). Wales eliminates national exams for many students. *Education Week*. Retrieved September 11, 2008, from http://www.edweek.org/ew/articles/2006/12/20/16wales.h26.html?qs=Wales.

Buchberger, F. & Buchberger, I. (2004). Problem solving capacity of a teacher education system as a condition of success? An analysis of the "Finnish case." In F. Buchberger & S. Berghammer (Eds.), *Education Policy Analysis in a Comparative Perspective,* (pp. 222-37). Linz: Trauner.

Chan, J.K., Kennedy, K.J., Yu, F.W., & Fok, P. (2008). Assessment policy in Hong Kong: Implementation issues for new forms of assessment. *The Hong Kong Institute of Education*. Retrieved September 12, 2008, from http://www.iaea.info/papers.aspx?id=68

Council for the Curriculum Examinations and Assessment. (2008a). Curriculum, Key Stage 3, Post-primary assessment. Retrieved September 12, 2008, from http://www.ccea.org.uk/

Council for the Curriculum Examinations and Assessment. (2008b). Qualifications. Retrieved September 12, 2008, from http://www.ccea.org.uk/

Dixon, Q. L. (2005). Bilingual education policy in Singapore: An analysis of its sociohistorical roots and current academic outcomes. *International Journal of Bilingual Education and Bilingualism, 8* (1), 25-47.

Dowling, M. (n.d.). Examining the exams. Retrieved September 14, 2008, from http://www.hkeaa.edu.hk/files/pdf/markdowling_e.pdf.

Eckstein, M.A. & Noah, H.J. (1993). *Secondary School Examinations: International Perspectives on Policies and Practice*. New Haven: Yale University Press.

European Commission. (2007/2008). The education system in Finland. *Eurybase, The Information Database on Education Systems in Europe*.

European Commission. (2006/2007). The education system in Sweden. *Eurybase, The Information Database on Education Systems in Europe*.

The Finnish Matriculation Examination. (2008). Retrieved September 8, 2008, from http://www.ylioppilastutkinto.fi/en/index.html

Finnish National Board of Education. (2007, November 12). Background for Finnish PISA success. Retrieved September 8, 2008, from http://www.oph.fi/english/SubPage.asp?path=447,65535,77331

Finnish National Board of Education. (2008, April 30). Teachers. Retrieved September 11, 2008, from http://www.oph.fi/english/page.asp?path=447,4699,84383

Finnish National Board of Education. (2008, June 10). Basic education. Retrieved September 11, 2008, from http://www.oph.fi/english/page.asp?path=447,4699,4847

Hong Kong Examinations Assessment Authority (HKEAA). (2007a). Introduction. *2007 Annual Report*. Retrieved September 14, 2008, from http://eant01.hkeaa.edu.hk/hkea/redirector.asp?p_direction=body&p_clickurl=http%3A%2F%2Fwww%2Ehkeaa%2Eedu%2Ehk%2Fen%2Fannual%5Freport%2Ehtml

HKEAA. (2007b, January 28). School-based assessment: Changing the assessment culture. http://www.hkeaa.edu.hk/en/hkdse/School_based_Assessment/SBA/

International Baccalaureate Organization. (2008). Diploma program assessment: Methods. Retrieved May 31, 2008, from http://www.ibo.org/diploma/assessment/methods/

International Baccalaureate Organization. (2006, May). IB Diploma Programme: Mathematics—Standard Level—Paper 2. Retrieved May 31, 2008, from http://www.ibo.org/diploma/curriculum/examples/samplepapers/documents/gp5_mathssl2.pdf

International Baccalaureate Organization. (2005, November). IB Diploma Programme: English A1—Higher Level—Paper 2. Retrieved May 31, 2008, from http://www.ibo.org/diploma/curriculum/examples/samplepapers/documents/gp1_englisha1hl2.pdf

Kaur, B. (2005). *Assessment of Mathematics in Singapore Schools—The Present and Future*. Singapore: National Institute of Education.

Kaftandjieva, F. & Takala, S. (2002). Relating the Finnish matriculation examination English test results to the CEF scales.

Korpela, Salla. (2004, December). The Finnish school—a source of skills and well-being: A day at Stromberg Lower Comprehensive School. Retrieved September 11, 2008, from http://virtual.finland.fi/netcomm/news/showarticle.asp?intNWSAID=30625

Laukkanen, R. (2008). Finnish strategy for high-level education for all. In N. C. Soguel & P. Jaccard (Eds.), *Governance and Performance of Education Systems*. Springer.

Lavonen, J. (2008). Reasons behind Finnish students' success in the PISA Scientific Literacy Assessment. University of Helsinki, Finland. Retrieved September 8, 2008, from http://www.oph.fi/info/finlandinpisastudies/conference2008/science_results_and_reasons.pdf

Murad, L.C. (2008). Hong Kong's education system: challenge for the future. Retrieved September 12, 2008, from http://www.lehigh.edu/~incntr/publications/perspectives/v20/Murad.pdf

Ng, P.T. (2008). Educational reform in Singapore: from quantity to quality. *Education Research on Policy and Practice*, 7, 5-15.

O'Donnell. (2004, December). Comparative tables and factual summaries—2004. *International Review of Curriculum and Assessment Frameworks*. Qualifications and Curriculum Authority and National Foundation for Educational Research. Retrieved September 11, 2008, from http://www.inca.org.uk/pdf/comparative.pdf

Petterson, A. (2008). The national tests and national assessment in Sweden. *Stockholm Institute for Education*. Stockholm, Sweden: PRIM gruppen. Retrieved on May 31, 2008, from http://www.prim.su.se/artiklar/pdf/Sw_test_ICME.pdf

Qualifications and Curriculum Authority. (2008a). Sweden: Assessment arrangements. Retrieved September 11, 2008, from http://www.inca.org.uk/690.html

Qualifications and Curriculum Authority. (2008b). England: Assessment arrangements. Retrieved May 27, 2008, from http://www.inca.org.uk/1315. http://education.qld.gov.au/corporate/newbasics/html/richtasks/richtasks.html_html

Qualifications and Curriculum Authority. (2009). Assessing pupils' progress: Assessment at the heart of learning. Retrieved May 23, 2009, from http://www.qca.org.uk/libraryAssets/media/12707_Assessing_Pupils_Progress_leaflet_-_web.pdf

Quality Assurance Division of the Education Bureau. (2008). Performance indicators for Hong Kong schools, 2008 with evidence of performance. Retrieved September 12, 2008, from http://www.edb.gov.hk/FileManager/EN/Content_6456/pi2008%20eng%205_5.pdf

Queensland Government. (2001). New basics: The why, what, how and when of rich tasks. Retrieved September 12, 2008, from http://education.qld.gov.au/corporate/newbasics/pdfs/richtasksbklet.pdf

Rustique-Forrester, E. (2005). Accountability and the pressures to exclude: A cautionary tale from England. *Education Policy Analysis Archives*. Retrieved October 1, 2009, from http://epaa.edu/epaa/v13n26

Schmidt, W. H., Wang, H.C. & McKnight, C. (2005). Curriculum coherence: An examination of U.S. mathematics and science content standards from an international perspective. *Journal of Curriculum Studies*, *37* (5), 525-59.

The Scottish Government. (2008). Schools: Attainment. Retrieved September 11, 2008, from http://www.scotland.gov.uk/Topics/Education/Schools/curriculum/Attainment

Scottish Qualifications Authority. (2004, March). Scotland's national qualifications: Quick guide. Retrieved September 11, 2008, from http://www.sqa.org.uk/files_ccc/NQQuickGuide.pdf

Singapore Examinations and Assessment Board. (2006). 2006 A-level examination. Singapore: Author.

Singapore Ministry of Education. (2007). Retrieved August 25, 2008, from http://www.moe.gov.sg/corpora/mission_statement.htm

State of Colorado House Bill 03-1108. (2003). Retrieved May 31, 2008, from http://www.state.co.us/gov_dir/leg_dir/olls/sl2003a/sl_153.htm

Swedish Institute. (1984, March). Primary and secondary education in Sweden. *Fact Sheets on Sweden*. Stockholm, Sweden.

Swedish National Agency for Education. (2005). The Swedish school system: Compulsory school. Retrieved May 31, 2008, from http://www.skolverket.se/sb/d/354/a/959

Texas Education Code, Section 51.968. (2005). Retrieved May 31, 2008, from http://www.legis.state.tx.us/tlodocs/79R/billtext/html/HB00130I.htm.

Valverde, G.A. & Schmidt, W.H. (2000). Greater expectations: Learning from other nations in the quest for 'world-class standards' in U.S. school mathematics and science. *Journal of Curriculum Studies*, *32* (5), 651-87.

Welsh Assembly Government. (2008a). Primary (3-11). Retrieved September 12, 2008, from http://old.accac.org.uk/eng/content.php?cID=5

Welsh Assembly Government. (2008b). Secondary (11-16). Retrieved September 12, 2008, from http://old.accac.org.uk/eng/content.php?cID=6

Appendix A:
Examples of International Assessment Systems

Country/ State	Description of core system	What kinds of assessments are used?	Who designs and grades assessments?
FINLAND	Student performance is evaluated on a sample basis by the Finnish education authorities at the end of second and ninth grades to inform curriculum and school investments. All other assessments are designed and managed locally, based on the national curriculum.	**National** Problems and written tasks that ask students to apply their thinking. **School-based** Research tasks, presentations, demonstrations.	**National** Designed by teachers through the Finnish Ministry of Education. Graded by teachers. **School-based** Teachers design and grade tasks based on recommended assessment criteria and benchmarks for each subject and grade within the national core curriculum.
	A voluntary matriculation examination is taken by most students to provide information to colleges. Students choose which subjects they will sit for (usually at least four), with the test in the students' mother tongue being compulsory.	The tests use mostly open-ended questions to evaluate skills including problem solving, analysis, and writing.	The exam is administered, organized and evaluated by The Matriculation Exam Board appointed by the Finnish Ministry of Education. Teachers grade the matriculation exams locally by using the official guidelines and samples of the grades are re-examined by professional raters hired by the Exam Board.
SWEDEN	Students take faculty-designed, nationally-approved examinations in year 9 and in the last two years of upper secondary school in Swedish, Swedish as a second language, English, and mathematics. Teachers use these assessments as one factor in determining students' grades at year 9, along with course grades and local assessments. In some cases, local regulations require schools to give an examination in year 5 in these same subjects. All other assessments are designed and managed locally. Teachers weight information from classroom work and assessments they design to determine whether students met the objectives of the national syllabus.	**National** Open-ended tasks requiring analysis of materials or problems, and written responses; materials may be given in advance of the test. **School-based** Coursework, research projects, diagnostic tasks, essays, problem sets.	**National** University faculty, with secondary teachers, design the tasks for the national exams at year 9 and the Upper Secondary level. Teachers grade the assessments during time set aside by regional authorities to calibrate grading practices to minimize variation across the region. They incorporate these scores into their grades for their courses. **School-based** Teachers design, administer, and grade tasks based on the national curriculum and syllabi, which outline "objectives to achieve" in each subject. Diagnostic materials for assessing syllabus goals are made available on an optional basis.

(continued on next page)

Appendix A: Examples of International Assessment Systems (cont'd)

Country/ State	Description of core system	What kinds of assessments are used?	Who designs and grades assessments?
ENGLAND	National curriculum assessments are enacted primarily as guidance for school-based formative and progress assessments conducted by teachers. A mandatory set of assessments at ages 7 and 11 includes externally developed tasks and observation scales implemented by teachers. Teachers choose which tasks and tests to use and when to use them, within certain parameters. Assessments for primary school are designed and managed locally, based on the national curriculum and guidance provided through the Assessing Pupils' Progress (APP) program.	**National** Observation scales completed by teachers regarding pupils' work and performance on specific kinds of tasks; written, oral, and performance tasks & tests. **School-based** Coursework, tests, projects, essays.	**National** The Qualifications and Curriculum Authority (QCA) develops the national assessments, which are scored by teachers, and a range of guidance and supports for in-school assessment. **School-based** Teachers evaluate student performance and work samples based on the national curriculum and syllabi. Extensive guidance for documenting pupil performance and progress, with indicators showing relationships to national standards, are provided through the Assessing Pupils' Progress project. Regional authorities support teacher training for assessment and in-school moderation.
	Most students voluntarily take a set of exams at year 11 (age 16) to achieve their General Certificate of Secondary Education (GCSE). If they take advanced courses, they may later take A-level exams, which provide information to universities. Students choose the exams they will take based on their interests and areas of expertise. About 75% of the exam grade is based on externally developed tests and 25% is school-based.	**National** Essays and open-ended problem solutions, oral language assessments. **School-based** Coursework, tests, projects.	**National** External exams are designed and graded by examining groups serving different schools (e.g. Oxford Cambridge, Ed Excel, the Assessments and Qualifications Alliance). **School-based** Teachers develop and score school-based components based on the syllabus.

(continued on next page)

Appendix A: Examples of International Assessment Systems (cont'd)

Country/ State	Description of core system	What kinds of assessments are used?	Who designs and grades assessments?
SINGAPORE	External examinations are given at the end of primary school (grade 6) in mathematics, science, English, and mother tongue (Malay, Chinese, or Tamil). Results are used to guide course placements in secondary school. All other assessments are school-based.	**National** Short and long open-ended responses. **School-based** Coursework, research projects, investigations.	**National** The Singapore Education Assessment Board designs the assessments and manages the assessment system. **School-based** Designed and graded by the classroom teacher in response to the syllabus.
	After four years of secondary school, students take the GCE N- or O-level examinations. Students choose the elective subject areas in which they want to be examined. Exams have school-based components that comprise up to 20% of the final score. Results are used as information for postsecondary education. GCE A-level examinations may be taken after two years of tertiary education.	**National** Short and long open-ended responses and multiple-choice items. **School-based** Research projects, laboratory investigations and experiments.	**National** The Singapore Education Assessment Board manages the assessment system. The GCE examinations are developed by the Cambridge International Examinations Group. **School-based** Teachers develop, implement, and score projects and other products that complement the external examinations.

(continued on next page)

Appendix A: Examples of International Assessment Systems (cont'd)

Country/ State	Description of core system	What kinds of assessments are used?	Who designs and grades assessments?
HONG KONG	The Hong Kong Territory-wide System Assessments occur at the equivalent of Grades 3, 6, and 9 in Chinese, English, and mathematics. The test is matrix-sampled, and results are reported to schools, but not publicly. Results are not reported for individual students. The goal is to inform curriculum planning within schools and to enable the government to assist schools that are struggling. An on-line bank of tasks is also available for teachers to use for diagnostic assessment of individual students.	**Territory-Wide** TSA items are written and oral open-ended items and tasks. **School-based** Essays, research projects, investigations.	**Territory-Wide** Assessments are developed and scored by the Hong Kong Education Examinations Authority. **School-based** Assessments are developed and scored by teachers. The Education Bureau encourages schools to develop multiple forms of assessment including projects, portfolios, and observations as well as tests, and looks for the variety of assessments in the performance indicators used for school evaluation.
	The Hong Kong Certificate of Education Examinations are taken at the end of secondary school to provide information to universities and employers. Students choose the areas they will sit for, beyond Chinese, English, mathematics, and liberal studies. These exams and the Diploma of Secondary Education that will replace them in 2012 include school-based assessments, comprising from 10% to 50% of the examination score.	**Territory-Wide** Written, speaking, and listening tasks **School-based** Projects, portfolios, lab experiments, reading activities, oral tasks	**Territory-Wide** Assessments are developed by teachers and higher education faculty, and scored by teachers who are trained as assessors. Tests are allocated randomly to scorers, and essay responses are typically rated by two independent scorers. **School-Based** School-based assessments are designed, administered, and scored by teachers in response to syllabus guidelines. Results are statistically moderated to ensure comparability within the province. The assessments are internationally benchmarked, through the evaluation of sample student papers, to peg the results to those in other countries.

(continued on next page)

Appendix A: Examples of International Assessment Systems (cont'd)

Country/State	Description of core system	What kinds of assessments are used?	Who designs and grades assessments?
AUSTRALIA	At the national level, a literacy and numeracy assessment is given at grades 3, 5, 7, and 9. States and localities manage their own assessment systems.	**National** Multiple-choice, short-answer, and extended written responses.	**National** Designed, administered, and scored by the Curriculum Corporation with questions and prompts contributed by state education agencies.
QUEENSLAND, AUSTRALIA	All additional assessments are school-based, developed by teachers based on the national curriculum guidelines and state syllabi. On an optional basis, schools may draw on a bank of "Rich Tasks" from the New Basics project that can be administered across grade levels and scored at the local level, with moderation.	**School-based** Open-ended papers, projects, and inquiries. Rich tasks are complex, interdisciplinary tasks requiring research, writing, and the development of multi-faceted products.	**School-based** Assessments are developed, administered, and scored by teachers. Scoring is moderated by regional panels of teachers and professors that examine scored portfolios of student work representing each score point from each grade level from each school. A state panel also looks at specimens across schools as well. Based on these moderation processes, schools are given instructions to adjust grades for comparability. Rich-tasks are developed by teachers with assessment developers; they are accompanied by scoring rubrics and moderation processes by which the quality of student work and scoring can be evaluated.

(continued on next page)

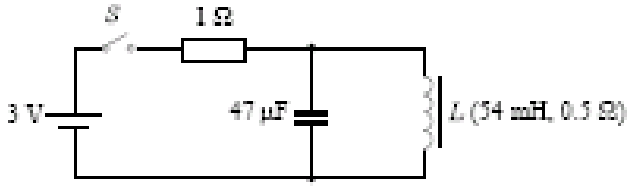
Appendix A: Examples of International Assessment Systems (cont'd)

Country/ State	Description of core system	What kinds of assessments are used?	Who designs and grades assessments?
VICTORIA, AUSTRALIA	All additional assessments are school-based until 11th and 12th grades, when students choose to take exams in different subject areas as part of the Vi**ctorian Certificate of Education (VCE), used to** provide information to universities and employers. The VCE exams have both external and school-based components. At least 50% of the total examination score is comprised of required classroom-based assignments and assessments given throughout the school year.	**State VCE** Multiple-choice (25%) and open-ended (75%) written, oral, and performance elements. **School-based** Lab experiments, essay, research papers and presentations.	The Victoria Curriculum and Assessment Authority (VCAA) establishes courses in a wide range of studies, oversees the development of the external examinations by teachers and university faculty, and ensures the quality of the school-assessed component of the VCE. Teachers score the open-ended items on the external exam and design and score the classroom-based assessments in response to syllabus guidelines. The quality of the tasks assigned by teachers, the work done by students, and the appropriateness of the grades and feedback given to students are audited through an inspection system, and schools are given feedback on all of these elements. In addition, the VCAA uses statistical moderation based on the external exam scores to ensure that the same assessment standards are applied to students across schools, adjusting the level and spread of each school's assessments to match that on the common exam.
INTERNATIONAL BACCALAUREATE	The International Baccalaureate (IB) Diploma Program, a program for students in grades 11 and 12 that is used worldwide, assesses students using school-based assessments throughout the two-year program and externally-developed exams at the end of the two-year program. School-based assessments comprise 20 to 50% of the examination score for each subject.	**External (IB-developed)** Essays, open-ended problem solutions, short answer, and multiple-choice items. **School-based** Speeches, projects, portfolio, presentations, investigations, labs, artistic performances.	**External** Designed, administered, and graded by trained IB examiners (usually current or former teachers). **School-based** Designed and graded by the classroom teacher based on a common syllabus and scoring criteria.

Appendix B: Hong Kong High School Physics Test

8. In Figure 8.1, a 47 μF capacitor, an inductor L and a 1 Ω resistor are connected with a cell of e.m.f. 3 V and negligible internal resistance. The inductor L is of inductance 54 mH and resistance 0.5 Ω. Initially the capacitor is uncharged.

Figure 8.1



(a) Find the current flowing in the 1 Ω resistor

(i) when the switch S is just closed ;

(ii) a few minutes after the switch S is closed.

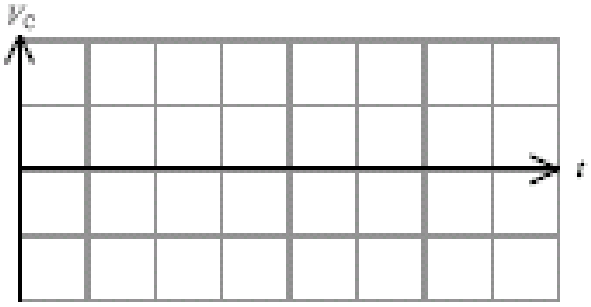
Explain briefly. (4 marks)

(b) (i) Calculate the maximum p.d. across the capacitor. (2 marks)

continued on next page

(ii) Find the energy stored in the inductor at the steady state. (2 marks)

(iii) If switch S is now opened, sketch the time variation of the p.d. V_C across the capacitor. (2 marks)



(c) State how you would modify the circuit so as to demonstrate that a large induced e.m.f. is produced across the inductor when switch S is suddenly opened. Explain briefly. (3 marks)

Linda Darling-Hammond, Co-Director
Stanford University Charles E. Ducommun Professor of Education

Prudence Carter, Co-Director
Stanford University Associate Professor of Education and (by courtesy) Sociology

Carol Campbell, Executive Director

Stanford Center for Opportunity Policy in Education
Barnum Center, 505 Lasuen Mall
Stanford, California 94305
Phone: 650.725.8600
scope@stanford.edu

http://edpolicy.stanford.edu