Benchmarking to International Best Practice

Integrating Standards, Curriculum, Assessments, and Teaching

The Process of Educational Reform

1983 – A Nation at Risk Report

"Our nation is at risk. Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world."

- National Commission on Excellence in Education

1989 – Goals 2000

"America will be first in the world in math and science by the year 2000."

—President George H.W. Bush & State Governors

2001 – No Child Left Behind

"We've spent billions of dollars with lousy results. Now it's time to spend billions of dollars and get good results."

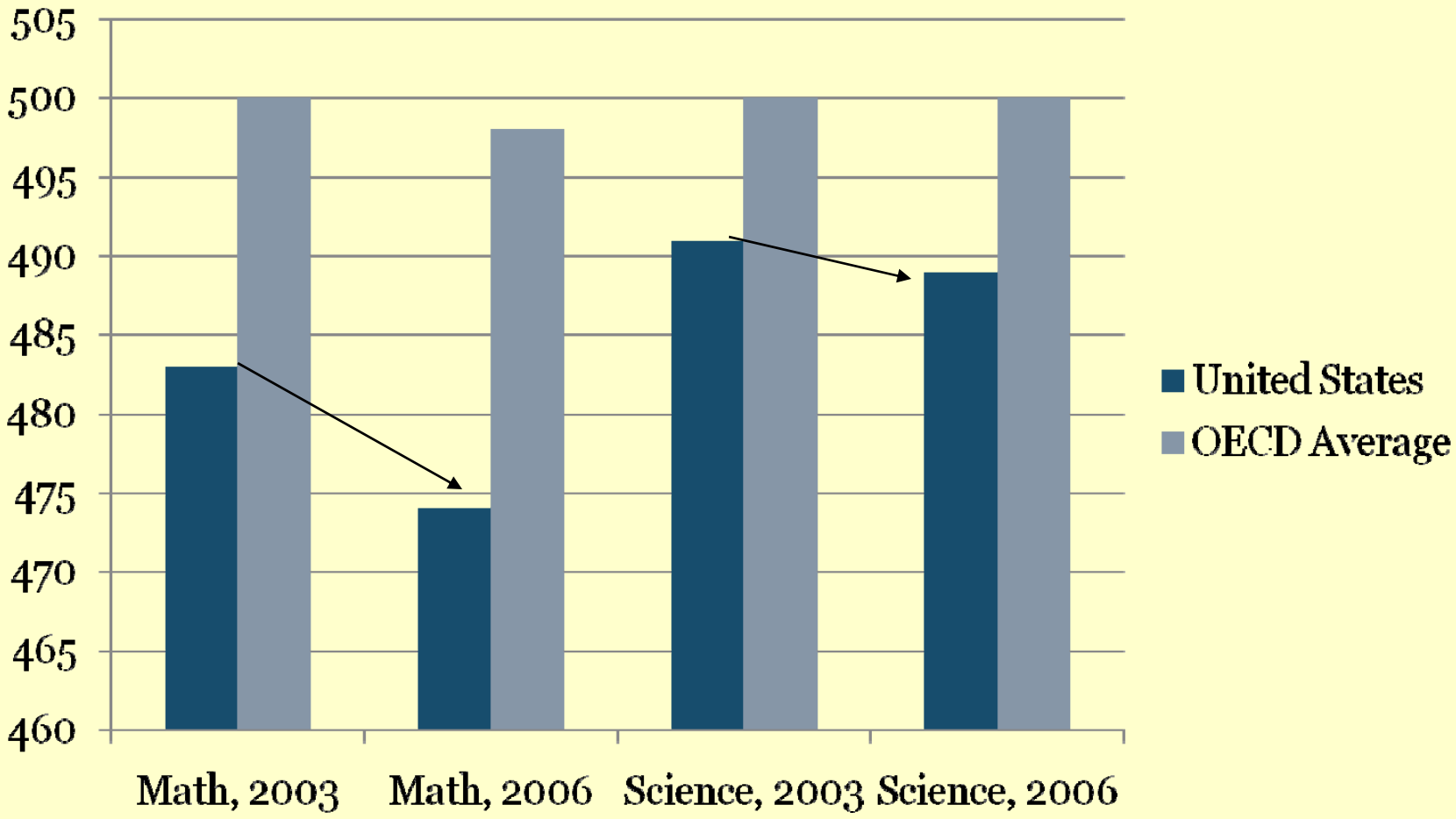
—President George W. Bush Jan. 9, 2002

International Outcomes

(8th Grade PISA Results in OECD Nations, 2006)

Science	Math
Finland	Finland
Canada	Korea
Japan	Netherlands
New Zealand	Switzerland
Australia	Canada
Netherlands	Japan
Korea	New Zealand
Germany	Belgium
United Kingdom	Australia
U.S. is # 29 / 40 top nations	U.S. is #35 / 40 top nations

U.S. Scores on PISA and PIRLS Have Dropped Since 2000



Differences Among Assessments

Most U.S. standardized tests are designed to assess if students learned what they were taught in school, focusing on recall and recognition of facts.

PISA is a set of international tests designed to assess if students can apply what they've learned to new problems and situations, focusing on inquiry and explanations of ideas.

Assessments in high-achieving nations increasingly emphasize demonstrations of learning authentic contexts.

Where we are Today

"I am calling on our nation's Governors and state education chiefs to develop standards and assessments that don't simply measure whether students can fill in a bubble on a test, but whether they possess 21st century skills like problem-solving and critical thinking, entrepreneurship and creativity."

-- President Barack Obama

March 10, 2009

Reform of Standards, Curriculum, and Assessment is Underway World Wide

…The goal is less dependence on rote learning, repetitive tests and a ‘one size fits all’ type of instruction, and more on engaged learning, discovery through experiences, differentiated teaching, the learning of life-long skills, and the building of character through innovative and effective teaching approaches and strategies…

-- Singapore Education Minister Tharman Shanmugaratnam, 2005

Singapore GCE A-Level Examinations

Time-based Written Papers

- 3 hour duration; 2 to 4 papers per H2 subject
- Open-ended essays, structured questions, case studies, source-based questions
- Externally set and marked by SEAB/CIE

School-based Coursework

- Longer duration of about 6 months
- Product (e.g. Artwork or design task), Oral Presentation, Independent Study
- Tasks set by SEAB/CIE, internally marked by teachers, externally moderated by SEAB/CIE)

SCHOOL-BASED SCIENCE PRACTICAL ASSESSMENT

To Assess Experimental Skills and Investigations, Students…

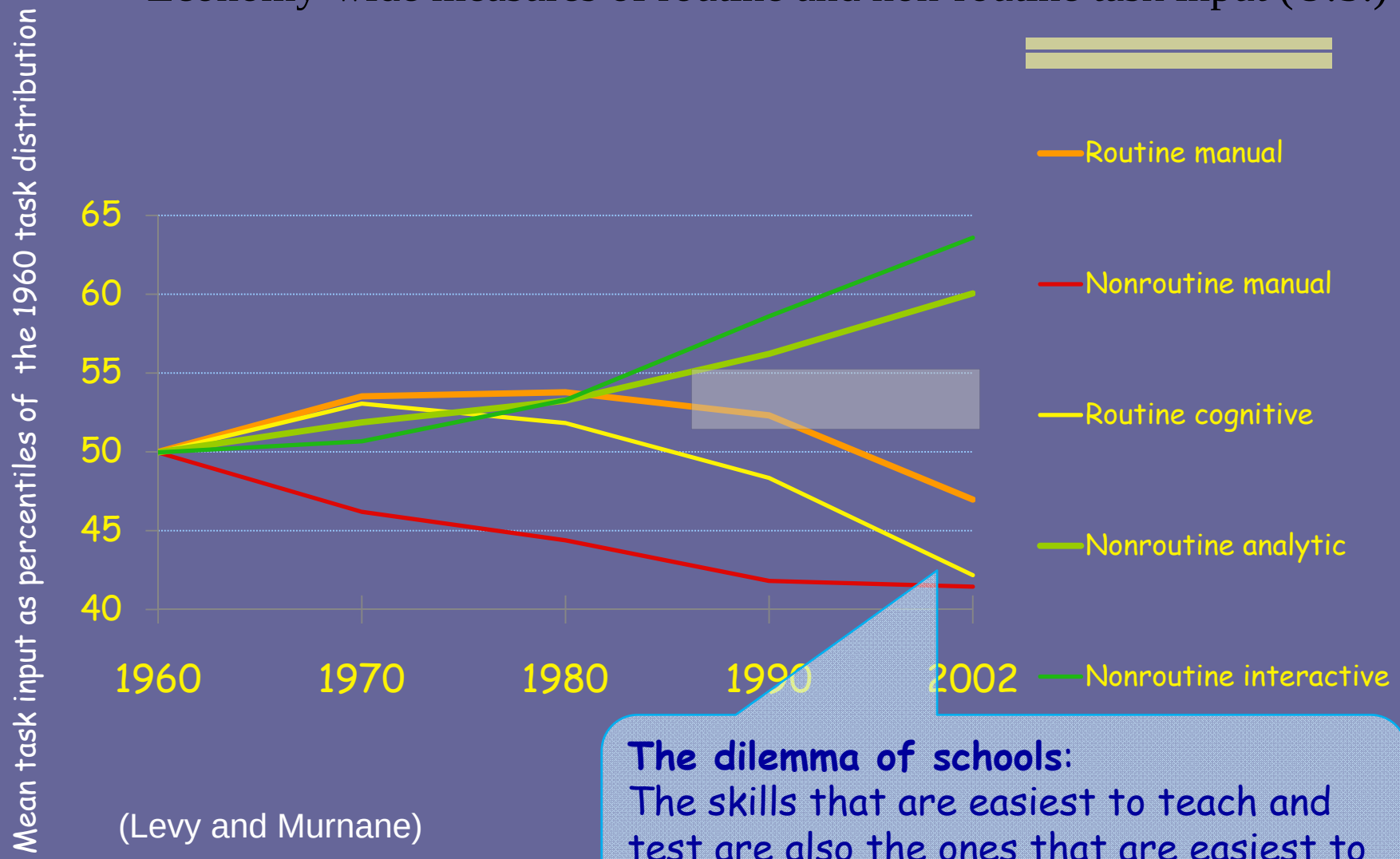
- Identify a problem, design and plan an investigation, evaluate their methods and techniques
- Follow instructions and use techniques, apparatus and materials safely and effectively
- Make and record observations, measurements, methods, and techniques with precision and accuracy
- Interpret and evaluate observations and experimental data

Project Work in Singapore, England, and International Baccalaureate

- Interdisciplinary coursework
- Extensive research (4000 word essay)
- Oral presentation
- Both Product and Process are assessed
- In Singapore, collaborative learning through group work is required and assessed

How the demand for skills has changed

Economy-wide measures of routine and non-routine task input (U.S.)



(Levy and Murnane)

The dilemma of schools:
The skills that are easiest to teach and test are also the ones that are easiest to digitize, automate, and outsource

Expectations for Learning are Changing

The new context means new expectations. Most studies include:

- Ability to communicate
- Adaptability to change
- Ability to work in teams
- Preparedness to solve problems
- Ability to analyse and conceptualise
- Ability to reflect on and improve performance
- Ability to manage oneself
- Ability to create, innovate and criticise
- Ability to engage in learning new things at all times
- Ability to cross specialist borders

NAEP, 8th and 12th Grade Science

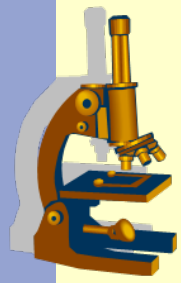
1. What two gases make up most of the Earth's atmosphere?

- A) Hydrogen and oxygen
- B) Hydrogen and nitrogen
- C) Oxygen and carbon dioxide
- D) Oxygen and nitrogen

2. Is a hamburger an example of stored energy? Explain why or why not.

_____

_____



A Rich Task: Science and Ethics Confer (Queenland, Australia)

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.

Applications of knowledge and skills assessed in *Science and Ethics Confer*

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

Worldwide, Reform Initiatives Generally Seek to . . .

- Emphasize expectations for higher-order skills along with rich content that represents core concepts and modes of inquiry.
- “Teach less, learn more” (Singapore): Focus the curriculum on standards that are fewer, higher, and deeper to allow more time to apply ideas in depth.
- Increase emphasis on project work and tasks requiring research, analysis, application, self-assessment, and production.
- Expand assessment of these intellectual skills, including the use of performance tasks with high levels of external validity.
- Develop assessments *of, as,* and *for* learning
- Arm teachers with learning progressions and greater capacity to use a wide range of assessment tools to analyze and support learning using school-based, as well as external, tests & tasks

England's GCSE
(General Certificate of Secondary Education)

1) End of course exams are open-ended short & long answer tests.
2) School-based ("controlled") assessments are taken under supervised conditions and are either set by the awarding body and marked by teachers or set by teachers and marked by the awarding body. These assessments count for:

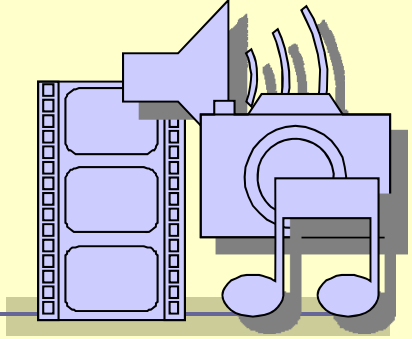
•25% of the exam score in: business studies, classical civilisation, English literature, geography, history, humanities, statistics, and

•60% of the exam score in: applied business, art and design, citizenship studies, construction and the built environment, dance, design and technology, drama, engineering, English, English Language, expressive arts, health and social care, home economics, hospitality and catering, ICT, leisure and tourism, manufacturing, media studies, modern foreign languages, music, physical education

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 2 options

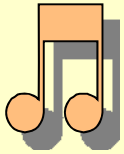
GCSE ICT Task (England)



Litchfield Promotions works with over 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 ICT 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.

Hong Kong's Rationale for Increasing School-Based Assessments

"The primary rationale for School Based Assessments (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, 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." (Hong Kong Education Examinations Authority, 2009).

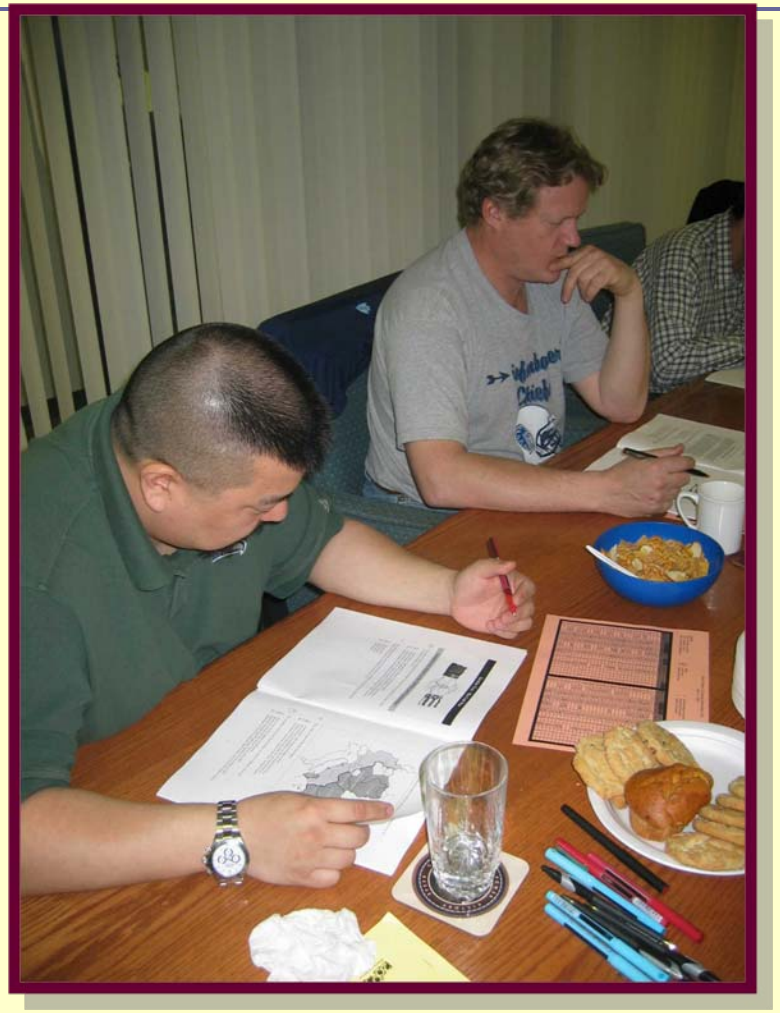
Common Practices Across Countries

- Assessments are part of a tightly integrated system of standards, curriculum, instruction, assessment, and teacher development at the state or national (in small countries) level
- Assessments include evidence of actual student performance on challenging tasks that evaluate 21st century standards of learning
- Teachers are integrally involved in the development and scoring of assessments (also college faculty)
- Assessments are used to provide information for improvement, not to determine school sanctions
- Assessments are designed to continuously improve teaching and learning

Teacher Collaboration: Test Design [from Alberta, Canada]

- Identify student characteristics
- Assist in exam blueprint development
- Ensure curricular “fit” of the exam
- Write and pilot prototype multiple-choice and written response forms
- Help develop writing assignments and their scoring criteria

Teacher Collaboration: Reviewing Tests



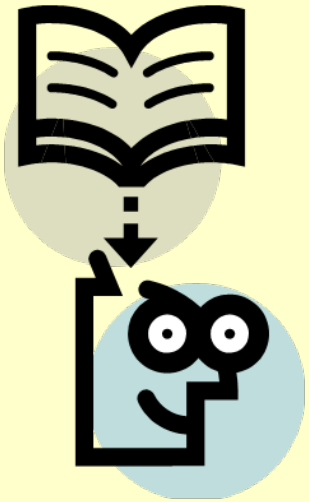
- Each new examination form is reviewed by a committee that includes classroom teachers.
- The committee examines both the written response and multiple-choice sections to ensure that the examination is fair, and demonstrates fidelity to the curriculum.

Teacher Collaboration: The Marking Process

- Teachers help select student work for use in setting benchmarks and training scorers.
- Teachers meet together to mark the written responses in a moderated process.
- Scores are reviewed, benchmarked, and calibrated reviewed to achieve consistency.

How Assessment Systems Aim to Improve Teaching and Learning

- Together, on-demand and curriculum-embedded assessments evaluate analytic and performance abilities that measure the full range of knowledge and skills represented in standards.
- Moderated teacher scoring of both components supports professional learning about assessment, standards, and teaching & more common instruction and grading
- Use of learning progressions to shape curriculum and assessments allows teachers to see where students are going and how to help them get there.
- School-based assessments provide models of good instruction and assessment, enhance curriculum equity, and allow teachers to see and evaluate student learning to inform teaching.



High School Biology Exam, Victoria, Australia

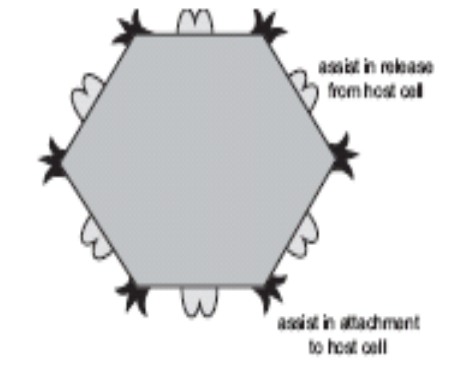
3. When scientists design drugs against infectious agents, the term “designed drug” is often used.

   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.

Analysis and Application of Knowledge

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.

Design and Scientific Inquiry

Before a drug is used on humans, it is usually tested on animals. In this case, the virus under investigation also infects mice.

C. Design an experiment, using mice, to test the effectiveness of the drug you have designed.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

School-Based “Coursework” Assessment

Victoria, Australia

In Unit 3 Biology, students are assessed on 6 pieces of work related the 3 outcomes specified in the syllabus.

- Outcome 1 – 3 practical tasks, one on plant & animal cells, another on enzymes, and a third on membranes.
- Outcome 2 – 2 practical activities related to maintaining a stable internal environment, one for animals, one for plants
- Outcome 3 – A research report / presentation on characteristics of pathogenic organisms and mechanisms by which organisms can defend against disease.

Creating Internationally Benchmarked Teaching and Learning Systems

- Common core standards / curriculum expectations
- Curriculum frameworks – lean but clear, including teaching and assessment guidance aimed at higher order skills
- State assessments at key junctures that include “on-demand” and curriculum-embedded components
- Teacher involvement in development and scoring
- Technology uses to enable richer, more targeted assessment
- Teacher preparation that includes extensive training in formative and summative performance assessment, knowledge of learning & development, content pedagogy, and teaching diverse students

Build on Successful Efforts

Consider what the highest-achieving states in the U.S. are doing:

-- MA, VT, ME, CT, NH, NJ

Use assessment knowledge that has been developed nationwide:

- NSF-supported performance assessments in math and science
- Science investigation tasks used as part of the New England Consortium (NECAP) and in Connecticut
- Writing and Math Portfolios in Vermont and Kentucky that were polished over time and scored reliably
- Performance tasks developed within and across content areas in Maine, Maryland, and Ohio
- New York's Revised Regents Exams that include extended essays and document-based questions
- Work samples used in Oregon's Certificates of Initial and Advanced Mastery
- Body of evidence assessments developed in WY, MT, MN
- Digital graduation portfolios incorporating student products and research developed in Rhode Island

Remember the Goal

- As Ted Sizer noted, the goal of education is for students to “learn to use their minds well” and to be able to apply what they know in the world beyond school
- Assessment *of, for,* and *as* learning should be designed with a primary aim of fostering these goals.
- Assessment systems should support the learning of everyone in the system: from students and teachers to school organizations and state agencies.

Costs

- **England**: Average cost per GCSE unit is 6.20 pounds ~ $12 - $15 per student per exam. Students choose at least 3 exams to take.
- **Alberta, Canada:** Average cost for set of 4 exams, graded by teachers in extensively moderated sessions, is $29/student.
- **Queensland:** Average cost per student for 11th and 12th grade moderated assessments in each course is ~ $18 - $20 per course.
- **U.S. Studies -** $9 to $30 per student depending on type of assessment and approach to scoring

Benefits

- Teacher involvement in scoring is actually a strong form of professional development leading to improved teaching
- Curriculum-embedded tasks enact standards and enhance curriculum equity for students
- Instead of “wasted time” or time away from instruction, curriculum-embedded assessments provide more time and focus for instruction while providing formative information about students

Uses of ICT-Based Scoring

(Hong Kong and International Baccalaureate)

- Examiners (teachers) trained and calibrated on-line
- Assessments increasingly conducted on-line
- Papers delivered on line for marking by trained teachers, with moderation
- Students' results tracked on-line and research managed with these results

Could be blended with some machine scoring of more limited open-ended tasks

U.S. NATIONAL HISTORY EXAMINATION

Who was president of the United States at the beginning of the Korean War?

A. John F. Kennedy
B. Franklin D. Roosevelt
C. Dwight Eisenhower
D. Harry Truman
E. Don't know

The Battle of the Bulge occurred during …

A. The Vietnam War
B. World War II
C. World War I
D. The Civil War
E. Don't know

Alberta Social Studies 30 Diploma Examination Questions

1. *A feature common to the Korean War and the Vietnam War was that in both conflicts:*

a) Soviet soldiers and equipment were tested against American soldiers and equipment

b) the United States became militarily involved because of a foreign policy of containment

c) the final result was a stalemate; neither side gained or lost significant territory

d) communist forces successfully unified a divided nation

2. *The Nazis' systematic genocide of ethnic and religious groups during the Second World War provided impetus for the:*

a) creation of a new autonomous state in Central Europe

b) establishment of a war crimes tribunal at Nuremberg

c) exclusion of Germany from the original membership of NATO

d) establishment of agencies fostering European economic and political cooperation

Example from Biology 30 Bulletin

January 2008 Biology 30 Diploma Examination Open-Response Question (Written Response 2), Sample Responses and Scoring Guide

Use the following information to answer the next question.

Autism is a complex behavioural disorder. The symptoms of autism vary greatly and occur in different combinations. Symptoms include a reduced ability to communicate, a reduced ability to develop relationships, difficulty coordinating facial muscles, and difficulty interpreting social cues.

In the late 1950s and early 1960s, the drug thalidomide was prescribed to pregnant women to combat morning sickness. Thalidomide was found to cause birth defects, such as stunted growth of the arms and legs. Some children also developed autism as a result of being exposed to thalidomide in utero. In comparison with the general population, the frequency of autism is many times higher in people with birth defects caused by thalidomide, which suggests that autism may originate early in embryonic development.

Time-Line of the Effects of Thalidomide on Embryonic Development

Age of embryo	20	21	22	23	24	25	26	27	28	29	30	31	32	33
	Nursing Mothers				Small ears and other malformations									
					Stunted arms									
							Stunted legs							

Scientists have genetically engineered mice that have symptoms similar to those of autism. These mice have a defective copy of the Hoxal gene, which is also present in humans, is normally active only during very early embryonic development.

Although people affected with autism are more likely to have the defective Hoxal gene than people without the disorder, the presence of the effective gene does not ensure the development of autism. Further investigation is required to determine whether enviromental factors work in conjunction with genes to produce autism.

Rodier, Patricia M. 2000. The early origins of Autism.
Scientific American, February, 56 -63

Time-Line of the Effects of Thalidomide on Embryonic Development - Continued

Identify two areas of the brain that can be affected in an individual with autism.

Explain the relationship between the areas of the brain identified and the symptoms of autism.

Identify one germ layer in which development could be disrupted by thalidomide and identify one structure that develops from this germ layer. Hypothesize how a person who has autism as a result of in utero exposure to thalidomide can have abnormal ear development but no malformations of the arms or legs.

Describe how the defective Hoxal protein is synthesized in the cytoplasm of a cell.
Explain how the defective Hoxal protein influences brain development and can lead to autism.

Identify and **describe** two technologies that can be used by scientists to replace an active Hoxal gene with a defective copy of the gene.

Example from Biology 30 Bulletin

Science

Score	*The student…* **Scoring Criteria**
5 - Excellent	• identifies two areas of the brain that are affected in an individual with autism and explains how the two brain areas identified relate to symptoms of autism • identifies a germ layer disrupted by thalidomide and one structure derived from the germ layer, and writes a complete hypothesis of how a person with thalidomide-induced autism can have abnormal ears but no limb malformations • describes translation and explains how the defective protein influences brain development and leads to autism
4 - Good	• identifies two areas of the brain that are affected and explains how one area that is identified relates to symptoms of autism • identifies a germ layer disrupted by thalidomide and one structure derived from the germ layer, and writes a partial hypothesis **or** identifies a germ layer that is disrupted by thalidomide and writes a complete hypothesis • describes translation and partially explains how the defective protein influences brain development to produce autism **or** partially describes translation and explains how the defective protein influences brain development
3 - Satisfactory	• identifies two areas of the brain that are affected in an individual with autism **or** identifies one area of the brain that is affected and explains how the area identified relates to the symptoms of autism • identifies a germ layer disrupted by thalidomide and the structure derived from the germ layer **or** identifies a germ layer disrupted by thalidomide and writes a partial hypothesis **or** writes a complete hypothesis • describes translation **or** explains how the defective protein influences brain development to produce autism **or** partially describes translation and partially explains how the defective protein influences brain development
2 - Limited	• identifies one area of the brain that is affected in an individual with autism • identifies a germ layer disrupted by thalidomide **or** writes a partial hypothesis • describes or explains one aspect of translation or how the defective protein influences brain development to produce autism
1 - Poor	• addresses only one of the three bullets at a 2 or a 3 level
0 - Insufficient	• does not address the question presented or provides an answer that is too brief to assess
NR	• does not provide a response