How Does the World Assess Student Learning?

Carol Campbell,

Stanford Center for Opportunity Policy in Education (SCOPE)

Asia Society Partnership for Global Learning 3rd Annual Conference

July 9, 2010

A New Day in Assessment

"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

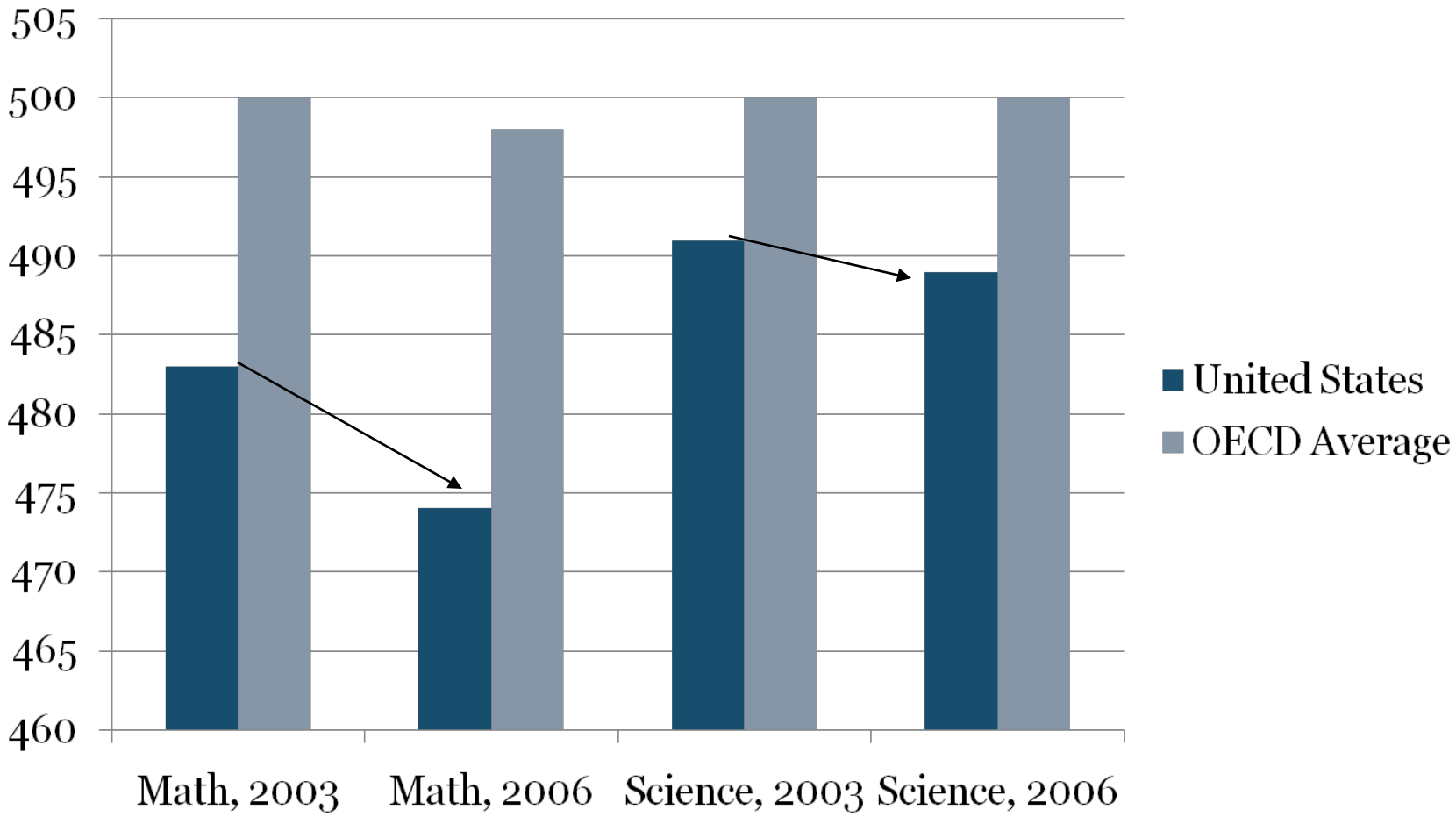
March 10, 2009

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

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



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**

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.

Shift from 'did students learn what we taught them?' to 'what can students do with what they have learned?'

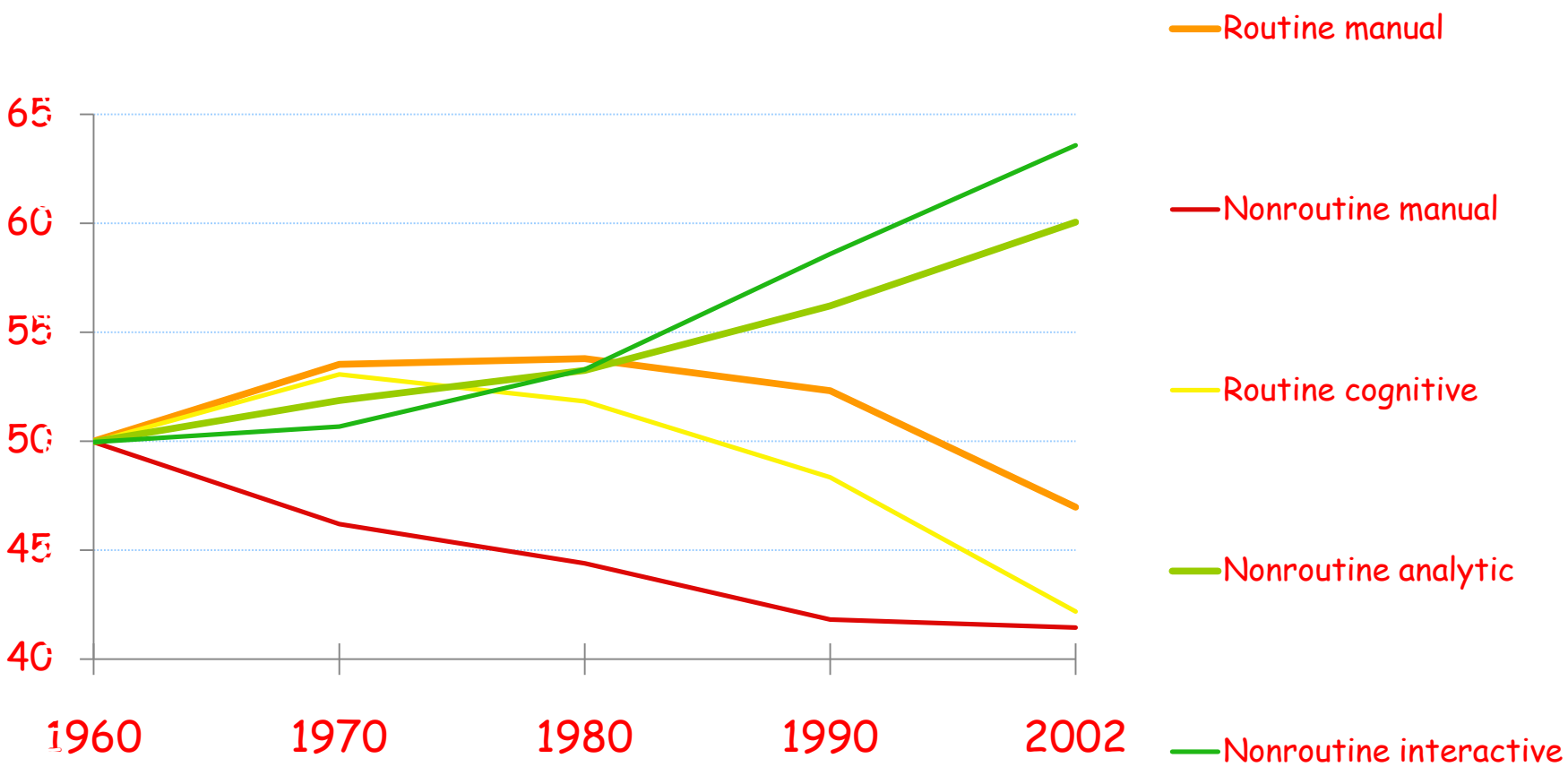
Assessments in high-achieving nations increasingly emphasize demonstrations of learning authentic contexts.

Two Kinds of Tests

- After a person learns to drive, he or she goes to the DMV to be tested. There, the person takes two different test to demonstrate his or her capability to be a licensed driver.
- One test is a multiple-choice test. The person must answer correctly questions about the laws related to driving.
- The other test is a performance test. The person must skillfully drive a car in a variety of road situations.
- **Which test is most important?**
- **Would you feel confident in knowing that a new driver had only been tested in one of these ways?**

How the demand for skills has changed

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



Expectations for Learning are Changing

The new context means new expectations. Most studies include:

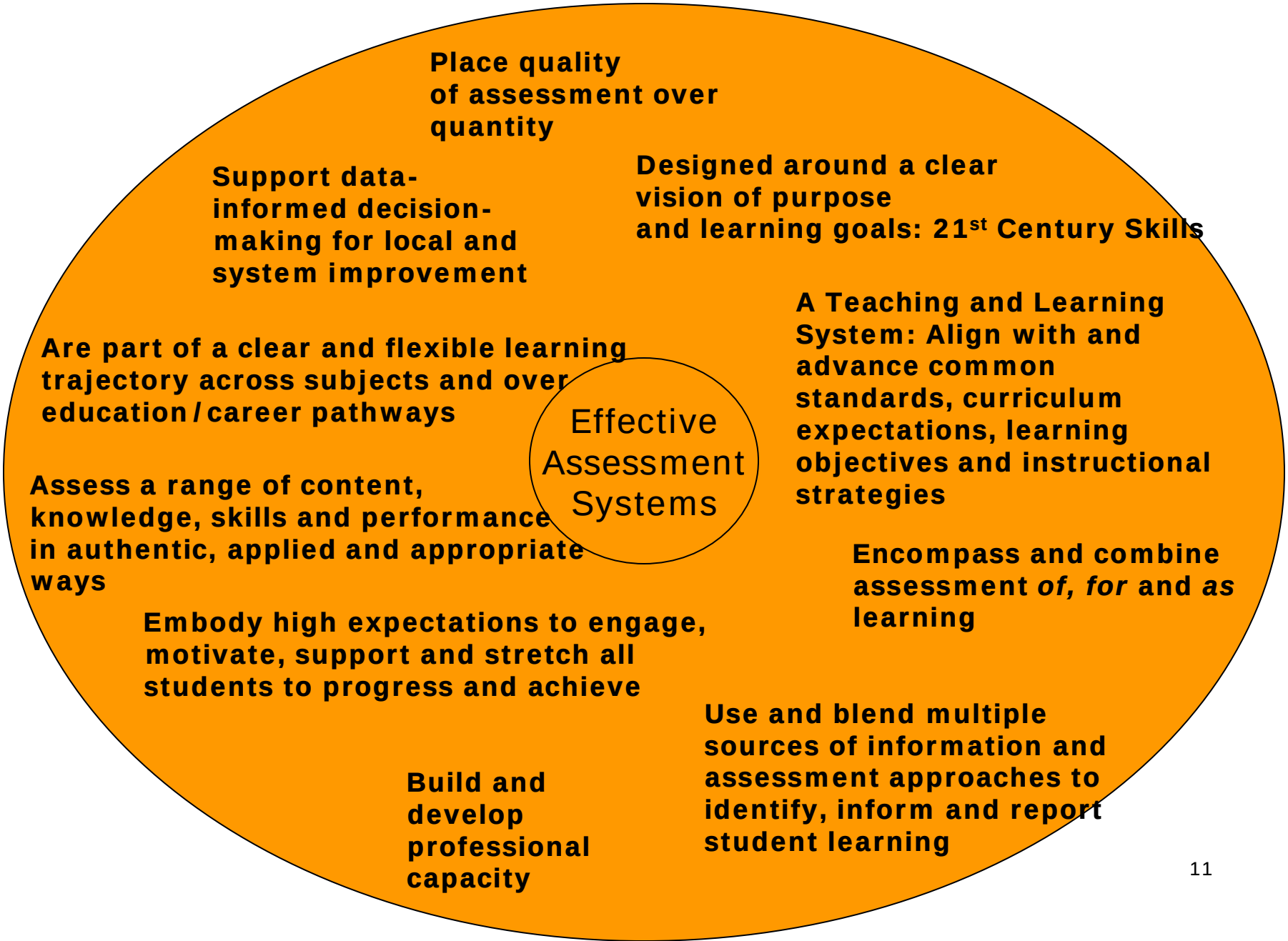
- Ability to communicate
- Adaptability to change
- Ability to work in teams
- Preparedness to solve problems
- Ability to analyse and conceptualise
- Ability to reflect on and improve performance
- Ability to manage oneself
- Ability to create, innovate and criticise
- Ability to engage in learning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

High-Achieving Countries Rely Increasingly on Performance Assessments

While multiple choice testing predominates in the U.S., most high-achieving countries largely use written & oral examinations, plus samples of student work, such as research projects and exhibitions, to evaluate what students have learned.

To "B" or
Not to "B"
……

Assessment Design Principles



1. Clear Vision of Purpose and Learning Goals - Specifying 21st century skills, college and career-readiness

High school examinations in England combine subject/curricular knowledge with:

Functional skills – practical skills in English, mathematics and information and communication technology" (ICT).

Personal learning and thinking skills (PLTS) – "require learners to be: Independent enquirers; Creative thinkers; Reflective learners; Team workers; Self-managers; Effective participants.

Singapore: **Life Skills and Knowledge Skills**

Example of Knowledge and Inquiry as a Humanities subject:

- Essay on theory and application of knowledge
- Critical thinking paper to analyze different information and arguments
- Independent study resulting in extended essay of 2,500 – 3,000 words

Purpose and learning goals - 21st Century Skills: Ontario - Achievement Chart – English, Grades 9-12 – Thinking Skills

Thinking – The use of critical and creative thinking skills and/or processes

Categories	50-59% (L1)	60-69% (L 2)	70-79% (L 3)	80-100% (L4)
	The student:			
Use of **planning skills** (e.g., generating ideas, gathering information, focusing research, organizing information)	uses planning skills with limited effectiveness	… some effectiveness	… considerable effectiveness	… high degrees of effectiveness
Use of **processing skills** (e.g. drawing inferences, interpreting, analysing, synthesizing, evaluating)	uses processing skills with limited effectiveness	… some effectiveness	… considerable effectiveness	… high degrees of effectiveness
Use of **critical/creative thinking processes** (e.g. oral discourse, research, critical analysis, critical literacy, metacognition, creative process)	uses critical/creative thinking processes with limited effectiveness	… some effectiveness	… considerable effectiveness	… high degrees of effectiveness

2. A Teaching and Learning System: Align with and advance common standards, curriculum expectations, learning objectives and instructional strategies

- Example: Ontario, Canada
- Content Standards
  - Curriculum for Grades 1-12 for all subjects and courses
    - Developed through extensive research, benchmarking, and consultation with all stakeholders
    - All teachers are required to teach the curriculum expectations
- Performance Standards (criterion-referenced)
  - Criteria are described for student achievement of the curriculum expectations for Grades 1-12
    - Four levels of achievement (provincial standard is Level 3, equivalent to 'B' or 70-79%
    - Four categories of knowledge and skills: Knowledge and Understanding; Thinking; Communication; Application
- Consistency:
  - Exemplars of student work for all curriculum
  - Standard provincial report cards - aligned with content and performance standards

3. Use and blend multiple sources of information and assessment approaches to identify, inform and report student learning

Multi-faceted assessments used internationally include:

- Assessments of knowledge (recall & analysis) and assessments of performance (demonstration of ability to apply knowledge in practice)
- Multiple-choice, constructed-response, extended tasks and projects
- On-demand and curriculum-embedded elements
- Externally-developed (by teachers and developers) and classroom-developed/managed
- Scoring made consistent through training, moderation, calibration, and auditing

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)

GRADE 12 GRADING (Diploma Subject), Alberta - Canada

Classroom 50%	Examination 50%
- Unit and Final Tests, - Quizzes - Projects (group/individual) - Labs - Portfolios - Oral Presentations - Exhibitions	Machine Scoreable 50% - 70% of items 50% of score weight
	Open – Ended 30% - 50% of items 50% of score weight

4. Embody high expectations to engage, motivate, support and stretch all students through assessing knowledge, skills and performance

For Example: Project Work in Singapore, England, and International Baccalaureate

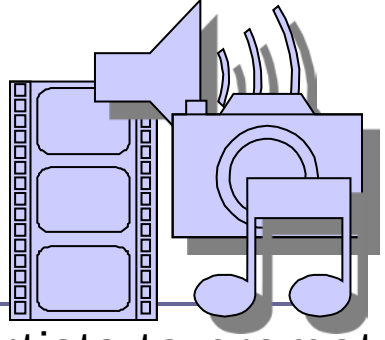
- Interdisciplinary coursework
- Extensive research (4000 word essay)
- Oral presentation
- Both product and process are assessed
- In Singapore, collaborative learning through group work is required and assessed

School-based science practical assessment, Singapore

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

GCSE ICT Task (England)



Litchfield Promotions works with over 40 bands and artists to promote their music and put on performances in England… Litchfield Promotions needs to create an ICT solution to ensure that they have all necessary information and that it is kept up to date…

Candidates will need to:

1) Work with others to **plan and carry out research** to investigate how similar companies have produced a solution….
2) 2) Clearly record and display your findings.
3) 3) Recommend a solution…
4) 4**) Produce a design brief…**

Produce a solution, ensuring that the following are addressed:

1) It can be modified to be used in a variety of situations.
2) 2) It has a friendly user interface.
3) 3) It is suitable for the target audience.
4) 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.**

Alberta Social Studies 30 Diploma Examination Questions

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, Alberta

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 the 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.

Describe three difficulties that researchers could encounter when they attempt to determine the cause of autism in humans.

5. Assess Range of Knowledge in Authentic, Applied Ways: English Language Arts -Constructing Scoring Rubrics: Identifying the Scoring Categories (Alberta)

Are students required/expected to:

Discuss the relationship among sources?

Demonstrate an understanding of multiple perspectives?

Apply learned concepts to new situations?

Demonstrate control of language conventions?

Interpret a source or sources?

Recall specific or factual knowledge?

Take and defend a position?

Demonstrate understandings of broad concepts?

Applying Knowledge: A Literacy Task (Grade 10, Ontario – Canada)

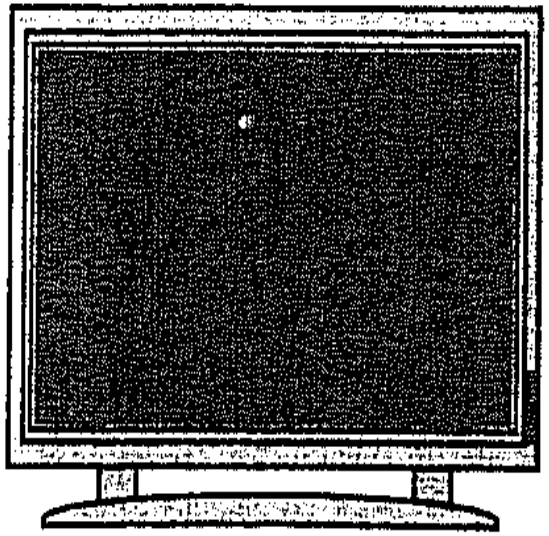
Task: Write a **minimum** of **three paragraphs** expressing an **opinion** on the topic below. Develop your main idea with supporting details (proof, facts, examples, etc.).

- **Purpose and Audience:** an adult who is interested in your opinion
- **Length:** The lined space provided for your written work indicates the approximate length of the writing expected.

Topic: Are today's famous people good role models for young people?

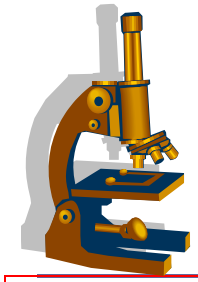
Applying Knowledge: A Mathematics Task (Ontario)

Sam is interested in buying a TV. At Fair Deal, the TV is regularly priced at $599.99 and is on sale for 20% off the regular price. At Big Big Discount, the same TV is regularly priced at $899.99 and is on sale for 30% off the regular price.



What is the difference in the sale price of the TV between these two stores?

Show your work.



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:

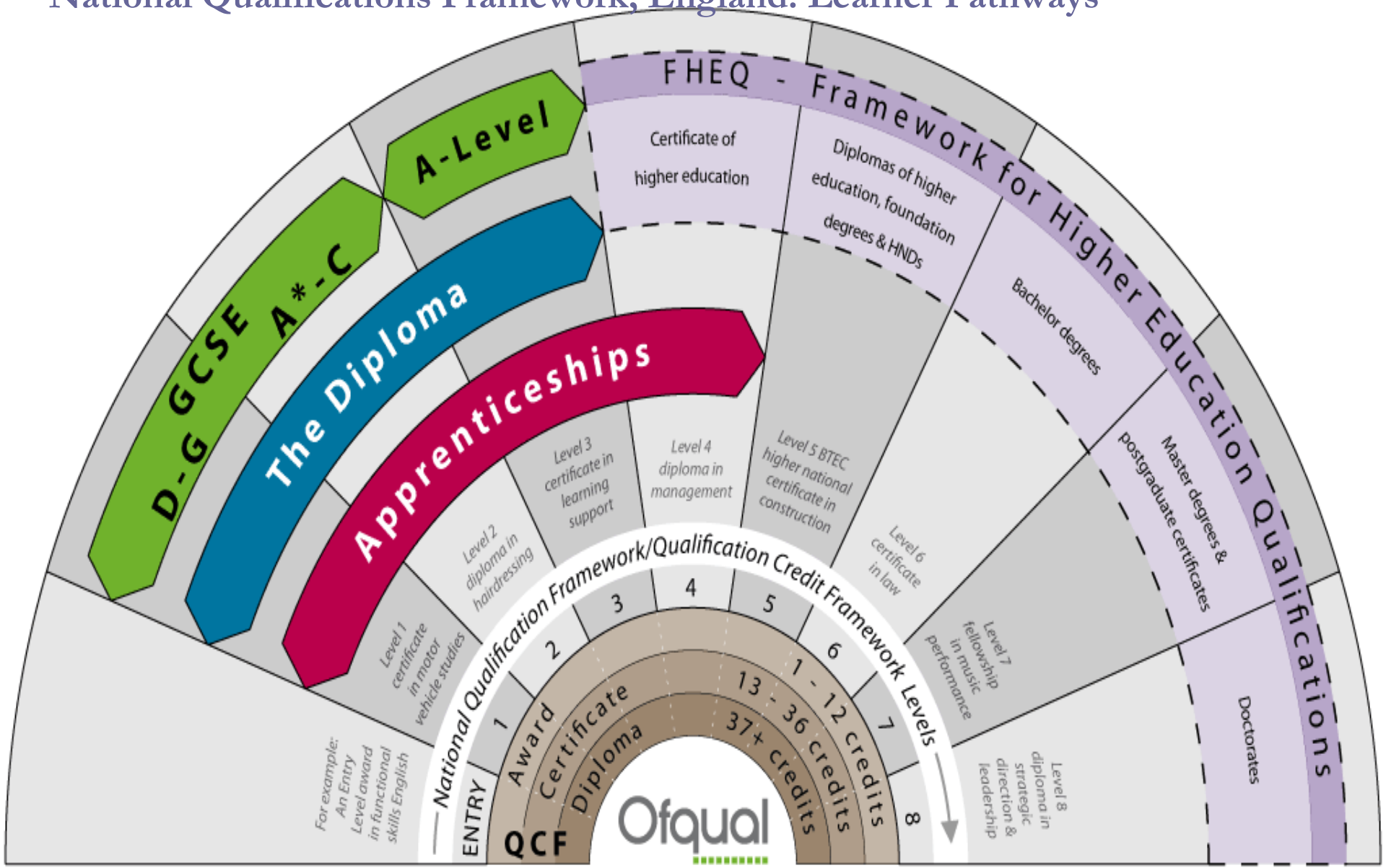
a) Provide a **written explanation of the fundamental technological differences** in some of the techniques used, or of potential use, in this area (included in the pre-conference package for delegates who are not necessarily experts in this area).

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.

6. Are part of a clear and flexible learning trajectory

National Qualifications Framework, England: Learner Pathways



7. Build and Develop Professional Capacity – Teachers are integrally involved

- Across higher performing countries, teachers are integrally involved in assessment design, validation, administration, moderation, and application for teaching and learning
- **Finland – teacher quality:**
  - teacher preparation includes strong focus on how to use formative performance assessment in service of student learning;
  - teachers use and adapt national curriculum and assessment criteria to develop more detailed curriculum, learning outcomes and assessment benchmarks
- **Australia (Queensland) - school-based assessment:**
  - Teachers develop, administer and score school-based assessments in relation to the national curriculum and state syllabi, and panels involving teachers and college faculty moderate assessments
- **Canada (Alberta, Ontario) – provincial/state assessments:**
  - Teachers integrally involved in: test design; item building; reviewing tests; confirming standards; marking process; examination advisory committee.

Professional Capacity: Test Design (Alberta, Canada)

- Example: Teacher collaboration on assessment design and marking, Alberta, Canada
- **Test design:**
  - Identify student characteristics
  - Assist in exam blueprint document
  - Ensure curricular “fit” of the exam
  - Pilot prototype multiple-choice and written response forms
  - Help develop writing assignments and their scoring

Item building:

- Item development sessions held throughout the province involve teachers in the creation of new multiple-choicce and written responses.

Teacher Collaboration – Reviewing Tests, Confirming Standards (Alberta)

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.

Confirming Standards:

- Before written responses are marked a committee of teachers meets with examination branch staff to select student work for use in marker training (example papers, training papers and reliability review papers).

Teacher Collaboration – Marking and Advising (Alberta)

The Marking Process:

- Each January and June approximately 1300 teachers meet to mark the written responses in a process with tight calibration to produce consistency in scoring.

Examination Advisory Committee:

- Once yearly representatives of various stakeholder groups, including the Alberta Teachers' Association meet with branch staff to review examination results from the previous school year, and to offer suggestions for the improvement of the examination program.

Scoring of Assessments - Ontario

- EQAO is an independent agency that coordinates all assessments and administers and reports on provincial large-scale assessments
- EQAO engages teachers in every step of the cycle of provincial assessments, from development and review to scoring (Approx. 2500 teachers per year).
  - Grade 9 Mathematics – raw score points = 46% multiple-choice, 54% open-response
  - Ontario Secondary School Literacy Test – raw score points = 48% multiple-choice, 52% open-response
  - Multiple choice items are machine scored. Written open-ended responses systematically scored by qualified scorers (mostly Ontario educators)
  - Use of generic rubric to describe levels of performance, item-specific rubric for each item, and anchors to illustrate descriptors for the score points to the rubric
  - Committee of educators define range of performance within each code of scoring rubric – 'range finding' development
  - Every scorer participates in extensive training. System of checks and monitoring applied throughout process
  - Quality assurance and statistical techniques to validate interrater reliability and scoring validity. Use of equating process to place student scores on common scale over two adjacent years and to determine cut scores

Teachers who engaged consistently in moderation processes were able to:

- Assess student performance more consistently, effectively, confidently, and fairly;
- Build common knowledge about curriculum expectations and levels of achievement;
- Identify strengths and areas for growth based on evidence of student learning;
- Adjust and acquire new learning by comparing one's thinking to that of another student or teacher;
- Share effective practices to meet the needs of all students, monitor progress, and celebrate growth.

Little, J.W. et al (2003). Looking at student work for teacher learning, teacher community and school reform. Phi Delta Kappan, 85(3), 185-192.

Key Roles in the Teacher Moderation Process

Principal's Role

- Schedules regular moderation sessions;
- Ensures resources are available for to interpret assessment results;
- Participates in moderation sessions, learning side-by-side with teachers and increasing shared knowledge about students' performance;
- Supports opportunities for distributed leadership.

Teacher's Role

- Collectively discusses results to plan and refine instruction
- Gives students feedback in a timely fashion to help improve performance;
- Actively participates in sessions through asking effective questions;
- Shares successful instructional strategies and resources with team, improving opportunities to support students' individual needs

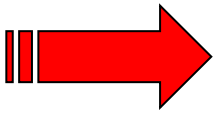
Student's Role

- Reflects on strengths and weaknesses in their learning;
- Incorporates and applies feedback to continuously improve work.

Adapted from Coalition of Essential Schools (CES) http://www.essentialschools.org/resources/60

Using Assessments to Inform Improvements: Example: Ontario

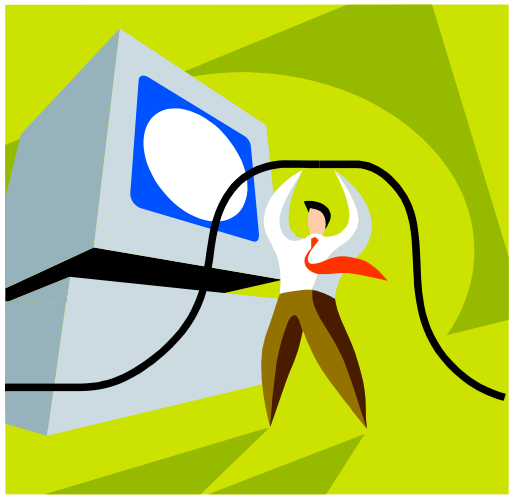
Teaching-Learning Critical Pathway.



- **Gathering Evidence**
- **Area of Greatest Need**
- **Current Practice**
- **Rubric, Data Wall, Culminating Task**
- **Professional Learning Community Action**
- **Moderated Marking**
- **Share Findings, Communication**

8. Support data-informed decision-making for local and system improvement

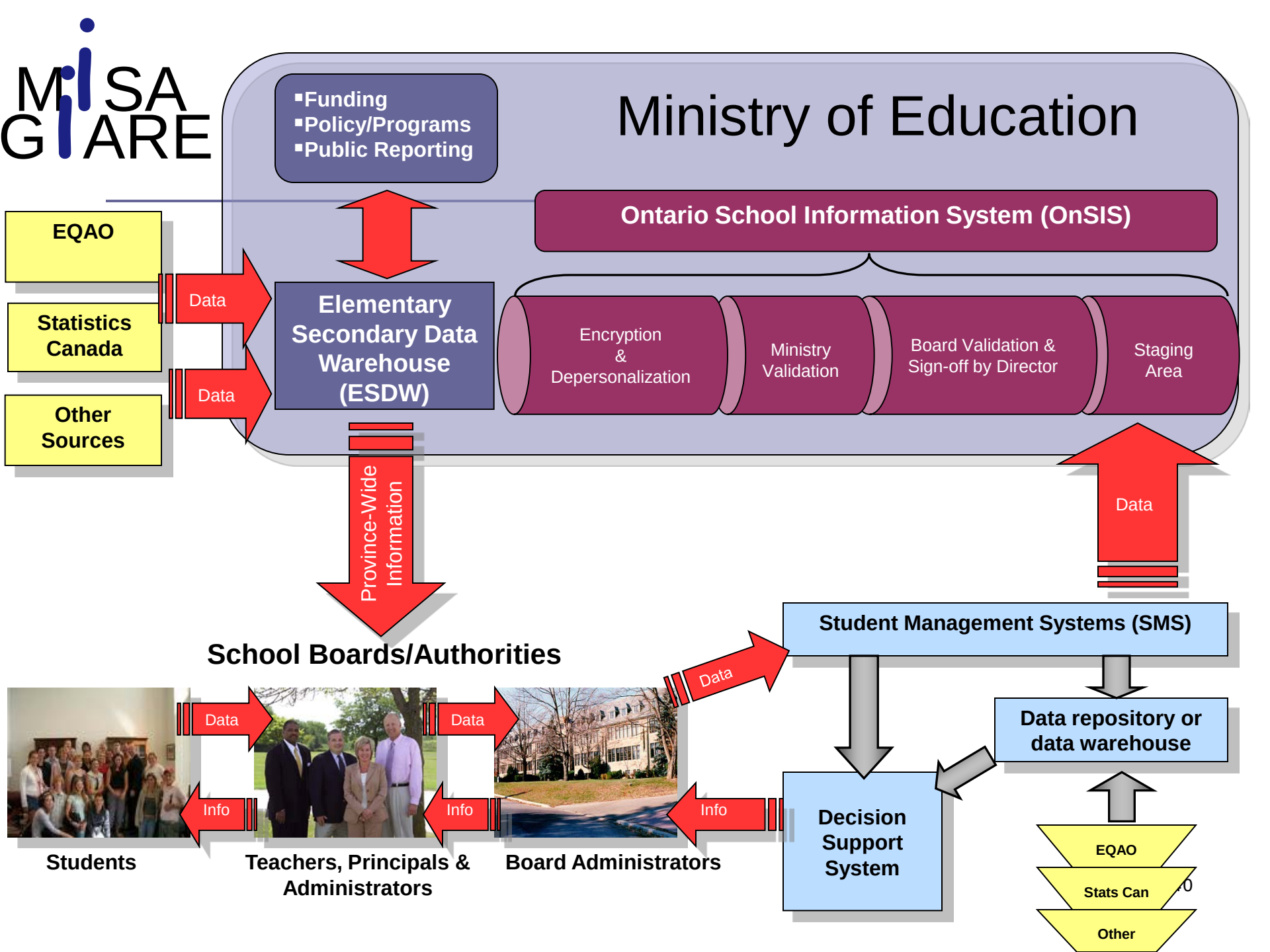
New technologies may support assessment quality and timeliness, as well as information systems that support reporting and accountability.



Technology Uses

- To distribute and administer assessments
- To enable simulations, research tasks, and other means for evaluating applied learning
- To support both human scoring and machine scoring of open-ended items

As a measure of the potential for technology to streamline performance testing, the National Assessment of Educational Progress has found that human and computer scoring of a set of physics simulations matches 96 percent of the time.


MISA GTA RE
Ministry of Education
Funding
Policy/Programs
Public Reporting
Ontario School Information System (OnSIS)
EQAO
Statistics Canada
Other Sources
Data
Data
Elementary Secondary Data Warehouse (ESDW)
Encryption & Depersonalization
Ministry Validation
Board Validation & Sign-off by Director
Staging Area
Data
Province-Wide Information
School Boards/Authorities
Student Management Systems (SMS)
Data
Data repository or data warehouse
Data
Data
Info
Info
Info
Decision Support System
Students
Teachers, Principals & Administrators
Board Administrators
EQAO
Stats Can
Other

9. Place quality of assessment over quantity

Examples:

- Australia – National standards (underway)
  State Syllabi, State and Local Assessments
  (3 grades plus high school)
- Canada – Provincial standards, syllabi & assessments
  (2 grades plus high school)
- England – National standards & curriculum
  School-based assessments + national tasks
  (1 grade plus high school)
  Five secondary examination boards
- Singapore – "Teach Less, Learn More"
  National standards and curriculum
  National exams + school-based tasks
  (1 grade plus high school)

10. Encompass and combine Assessment *of, for* and *as* learning

Assessment…

Of learning - summative assessments to confirm student knowledge, skills, understanding of curriculum, proficiency

For learning – formative assessments using information throughout the learning process to investigate student learning, inform instruction and provide feedback to support student progress

As learning – assessment as a process of meta-cognition involving student self-assessment and understanding of own learning processes and knowledge

Hong Kong: Learning to Learn and Use of 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 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 more reliable assessment of each student… SBA is to promote a positive impact on teaching and learning**. It can serve to motivate students by engaging them in meaningful activities; and for teachers, it can reinforce curriculum aims and good teaching practices…"

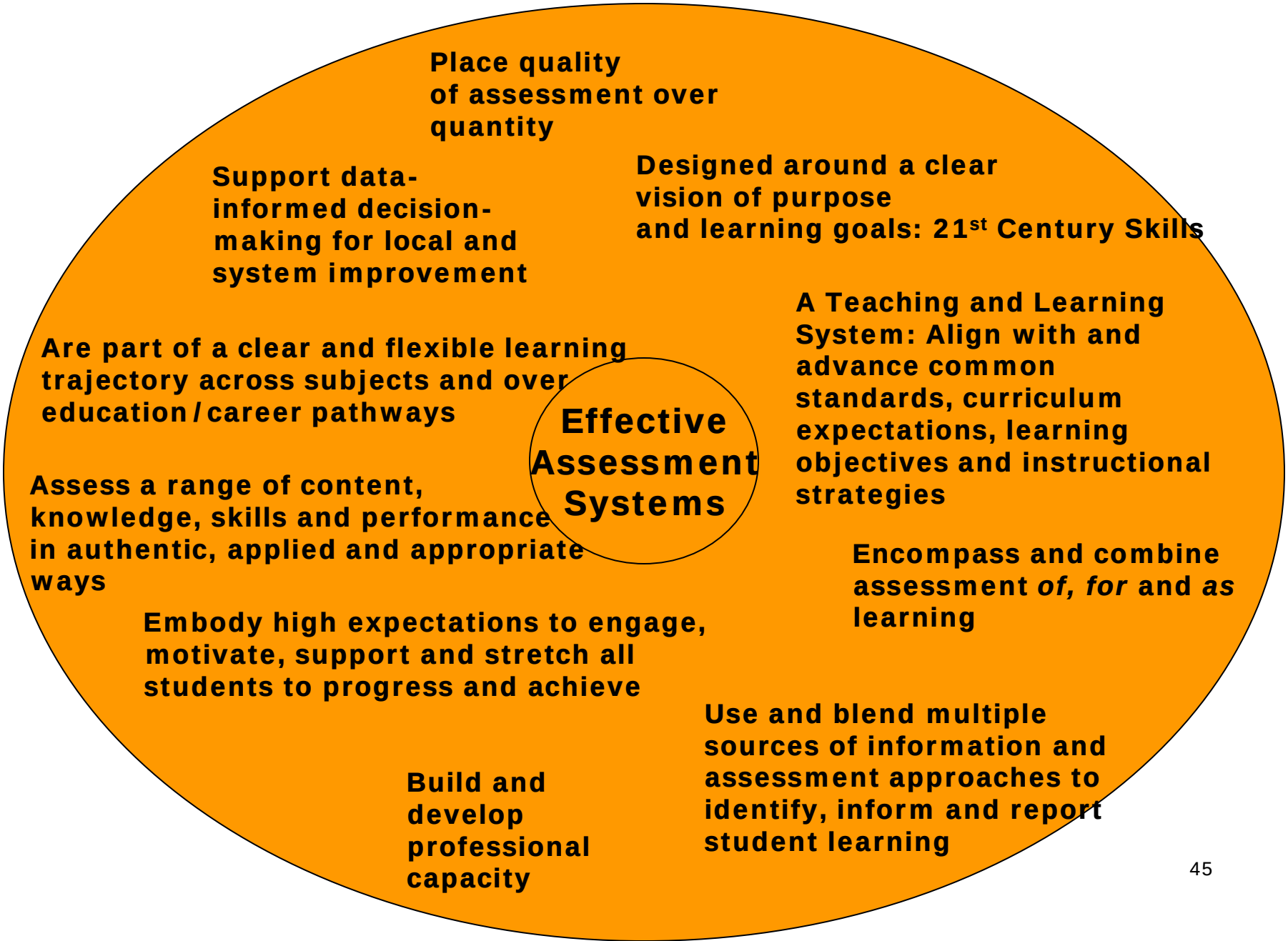
(Hong Kong Education Examinations Authority, 2009).

"If you want to appear accountable, test your students.

If you want to improve schools, teach teachers to assess their students.

If you want to maximize learning, teach students to assess themselves."

Assessment Design Principles



Race To The Top Assessment Grants

$350 million for state consortia to develop new assessments "that are valid, support and inform instruction, provide accurate information about what students know and can do, and measure student achievement against standards designed to ensure that all students gain the knowledge and skills needed to succeed in college and the workplace. These assessments are intended to play a critical role in educational systems: provide administrators, educators, parents, and students with the data and information needed to continuously improve teaching and learning; and help meet the President's goal of restoring, by 2020, the nation's position as the world leader in college graduates" (Federal Register/Vol.75, No.68/ April, 9, 2010)

2 categories:

- Comprehensive Assessments Systems grants
- High School Course Assessment Programs grants

How Might U.S. Assessment Systems Become Internationally Comparable?

- Develop a **coherent**, **aligned and integrated system** that links curriculum, assessments, instruction, and student opportunities-to-learn
- Create lean, but clear **curriculum guidance** (frameworks or syllabi) around the new Common Core standards that can guide:
  - **End-of-year summative** assessments built on well-articulated content and performance standards
  - **During-the-year performance tasks** embedded in curriculum and scored as part of the overall assessment system
  - **Formative assessment** supports that provide rich information
  - Focused **preparation and professional development** linked to teachers' roles in developing and scoring assessments, as well as instruction

SCOPE: Commissioned Papers

http://edpolicy.stanford.edu

Suzanne Lane, *Performance Assessment: The State of the Art.*

Raymond Pecheone and Stuart Kahl, *Developing Performance Assessments: Lessons from the United States.*

Brian Stecher, *Performance Assessment in an Era of Standards-Based Educational Accountability.*

Jamal Abedi, *Performance Assessments for English Language Learners.*

Linda Darling-Hammond, with Laura Wentworth, *Benchmarking Learning Systems: Student Performance Assessment in International Context.*

Lawrence Picus, Will Montague, Frank Adamson, and Maggie Owens, *A New Conceptual Framework for Analyzing the Costs of Performance Assessment.*

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.*

FUNDED BY FORD AND NELLIE MAE EDUCATION FOUNDATIONS

Thank You!

CarolCampbell@stanford.edu

http://edpolicy.stanford.edu