Technologies in Large-Scale Assessments
New Directions, Challenges & Opportunities

Michal Beller
Director-General of RAMA
The National Authority for Measurement and Evaluation in Education, Israel

International Large-Scale Assessment Conference
March 16-18, 2011
ETS, Princeton, NJ
Personal Preface

The Hebrew University of Jerusalem

The Open University of Israel

Providing research to drive innovation in assessment.

RAMA - The National Authority for Measurement and Evaluation in Education
In this Presentation

Description of some recent developments of technology-based Large Scale Assessments (LSA) & International LSA (ILSA), and address the following questions:

- Whether the computerized revolution of I/LSA will be merely a technological leap forward, or serve as a catalyst for a more profound pedagogical change in the way instruction, learning and assessments will be conducted in the next era?
- Will it foster the integration of 21st century competencies and expertise into teaching and instruction of all content areas?
- Will it facilitate the creation of new methodologies for a better use of technology and assessment in the service of learning?
Introduction

- Assessment serves a critical role in education as it holds education systems accountable, and at the same time it is a gateway to any systemic change.

- Assessment matters because it drives change at all levels, from classroom to legislature, though its effects range from the positive to the negative.

- Therefore, the assessment and the mode of assessment should be designed carefully to ensure a positive impact on education.

- For this to happen, assessments should be more complete, more authentic and fully integrated into the learning and teaching process.

- Reaching such goal is a difficult endeavor that can be enhanced by technology.
Technology is not a Goal in Itself

Extracting the full learning return from a technology investment requires the triangulation of:

- content
- sound principles of learning
- high-quality teaching

All of which must be aligned with assessment and accountability

The power of this technology for learning math is unleashed when the tools are:

- used long-term
- are integral to the instruction
- not just computational tools
- and...

Are used in both instruction and assessment activities.
Assessment is not a Goal in Itself

Full Alignment is Required

Introduction

CET

Large Scale Assessments (LSA)

Content & Performance Standards

Teaching & Learning

Curriculum

Assessment

Recent Developments

International Large Scale Assessments (ILSA)

The Challenges
Negative Consequences of High Stakes Assessments

Teaching & Learning

Content & Performance Standards

"Teaching to the Test"
Curriculum
Diverting Resources

Assessment

Introduction
CET
Large Scale Assessments (LSA)
International Large Scale Assessments (ILSA)
Recent Developments
The Challenges
Effective integration of technology in the learning and teaching process is a common challenge to almost every education system in the world.

Effective integration of assessment in learning and teaching is an even greater challenge, particularly due to the need to strike the right balance between formative and summative assessments.
Effective Integration of Technologies

Learning Technologies

Learning & Teaching

Assessment Technologies

Summative

Formative
Roles of Technology in Assessment

Improving efficiency and Quality of existing practices

- Test Creation Assistant
- Computer delivery of tests (CAT)
- Automated scoring of complex items (e-Rater)

Expanding the domain of testing

- Simulated-based assessment tasks
- Intelligent tutoring systems
- Virtual reality systems
New Era of Integrated, Learning-centered Assessment Systems

- Large-scale testing of educational outcomes benefits already from technological applications that address logistics (such as development, administration, and scoring of tests, as well as reporting of results).

- Innovative applications of technology also provide rich, authentic tasks that challenge the sorts of integrated knowledge, critical thinking, and problem solving seldom addressed in paper-based tests.
New Era of Integrated, Learning-centered Assessment Systems

- Such tasks can be used on both large-scale and classroom-based assessments.

- Balanced assessment systems can be developed that integrate curriculum, formative and summative assessments across classroom, district, state, national, and international levels.

- The goal is to utilize the potential of technology to launch a new era of integrated, learning-centered assessment systems.

Quellmalz and Pellegrino, 2009
Universal Design of Assessment

- Technology allows the development of assessments using Universal Design principles that make assessments more accessible, effective, and valid for students with greater diversity in terms of disabilities and limited language proficiency.

- Main design principles:
  - Alternative representations of same concepts and skills
  - Presenting information through multiple modalities
  - Assistive technologies
Framework for 21st Century Learning

Five white papers:

- Defining 21st century skills
- Perspectives on methodological issues
- Technological issues for computer-based assessment
- New assessments and environments for knowledge building
- Policy frameworks for new assessments
Working Group on Technological Issues

Led by Benő Csapó the group looks at the many ways technology can be used to improve assessment:

- Detecting and recording students’ psychomotor, cognitive and emotional characteristics
- Enabling adaptive or personalized testing, in which students are presented with a unique set of tasks focused on their individual performance levels
- Administering dynamic problems that change the conditions, information or instructions as the student is working
- Evaluating how fast students arrived at the answers and the processes they used - making the students’ thinking visible
Requirements for Effective Integration of Technologies and Assessment

Assessment should be:

- **Aligned** with the development of significant 21st century learning goals
- **Adaptable and responsive** to new developments in technology and in education
- **Largely performance-based**
- **Add value** for teaching and learning by providing information that can be acted upon by students, teachers, and administrators
- **Better meet general criteria for good assessment** (i.e., be fair, technically sound, valid for purpose, and part of a comprehensive and well-aligned system of assessment at all levels of education)
Goals of Integration of Technology into Assessment

Thougtful integration of technology into assessment may serve two complementary goals:

- An innovative use of technology for learning and for the assessment of complex cognitive skills (e.g., Diagnoser, SimScientists, CET).

- A new accountability paradigm which fully integrates sequential formative assessments and periodic summative assessments via computers (e.g., CBAL to be discussed later).
The SimScientists program in WestEd's Science, Technology, Engineering & Mathematics (STEM) program is comprised of a portfolio of research and development projects that focus on the roles that simulations can play in enriching science learning and assessment.

The capabilities of technology allow modifications of simulation-based activities to offer accommodations for English learners and students with disabilities. Science simulations can be used in curriculum activities, as embedded, formative assessments, and as summative assessments.

The SimScientists program investigates design principles to guide the creation of effective simulations as learning and assessment tools.

In addition, the SimScientists program studies how science simulations can be used at different levels of the educational system - classroom, district, and state in balanced state science assessment systems.
Introduction

Large Scale Assessments (LSA)

Recent Developments

The Challenges

For further details - click here
The Role of Technology in National Large Scale Assessments (LSA)
Organizational and Technical Issues

- In large-scale implementation efforts up until now, the school infrastructure are usually given, and either the existing equipment is used as is they are, or new equipment is installed for assessment. However, this presentation will not deal with technological opportunities ahead of us, nor with technological barriers existing today (e.g., access to computers in schools, bi-directional adaptation of the test platform).

- As well as future technologies, hold the potential to dramatically change the way assessment will be implemented in the future.
Examples of Leading Computer-Based National Assessments

In Iceland, projects are underway to change the current paper-and-pencil national testing “platform” to a computer-based platform that will be delivered over the Internet.

In Denmark, the full range of national student assessments has been recently migrated to a computer-based platform.

This year Australia will administer the Australian National Assessment Program 2011 Information and Communication Technology Literacy Assessment.
Research Projects:

- MOL - Math Online
- WOL - Writing Online
- TRE - Problem Solving in Technology-Rich Environment

Large Scale Implementations

- 8th Grade
- 12th Grade
- TELA 2014
Problem 1

How do different payload masses affect the altitude of a helium balloon?

Design Experiment

Run Experiment

Interpret results

<table>
<thead>
<tr>
<th>Payload Mass (pounds)</th>
<th>Amount of Helium (cubic feet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2275</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Altitude (feet)</th>
<th>Balloon Volume (cubic feet)</th>
<th>Time to Final Altitude (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2275</td>
<td>0</td>
</tr>
</tbody>
</table>
Problem 1

How do different payload masses affect the altitude of a helium balloon?

- Design Experiment
- Run Experiment
- Interpret results

<table>
<thead>
<tr>
<th>Payload Mass (pounds)</th>
<th>Amount of Helium (cubic feet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>2275</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Altitude (feet)</th>
<th>Balloon Volume (cubic feet)</th>
<th>Time to Final Altitude (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12174</td>
<td>3083</td>
<td>122</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>final altitude (feet)</th>
<th>payload mass (pounds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>0</td>
</tr>
<tr>
<td>35000</td>
<td>10</td>
</tr>
<tr>
<td>30000</td>
<td>20</td>
</tr>
<tr>
<td>25000</td>
<td>30</td>
</tr>
<tr>
<td>20000</td>
<td>40</td>
</tr>
<tr>
<td>15000</td>
<td>50</td>
</tr>
<tr>
<td>10000</td>
<td>60</td>
</tr>
<tr>
<td>5000</td>
<td>70</td>
</tr>
<tr>
<td>0</td>
<td>80</td>
</tr>
</tbody>
</table>

© 2011 # 26
Introduction

CET

Large Scale Assessments (LSA)

International Large Scale Assessments (ILSA)

Recent Developments

The Challenges

The framework specifies three kinds of practices that students are expected to demonstrate when responding to test questions:

- **Understanding Technological Principles** focuses on how well students are able to make use of their knowledge about technology.

- **Developing Solutions and Achieving Goals** refers to students’ systematic use of technological knowledge, tools, and skills to solve problems and achieve goals presented in realistic contexts.

- **Communicating and Collaborating** concerns how well students are able to use contemporary technologies to communicate for a variety of purposes and in a variety of ways, working individually or in teams, with peers and experts.

The assessment will be completely computer-based.
Oregon (OR)

OAKS - Oregon Assessment of Knowledge & Skills, is a new online testing system assessing students' mastery of Oregon content standards. The OAKS online operational test is now available for Reading, Mathematics, Science, and Social Sciences.

West Virginia (WV)

techSteps is a personalized, project-based technology literacy curriculum that infuses technology skills into core instruction, promoting core subject area outcomes while also teaching skills for 21st century.
The Role of Technology in International Large Scale Assessments (ILSA)
The growing popularity of International Large Scale Assessments (ILSA) has been accompanied, along with its positive impact, by the following:

- In many countries today ILSA frameworks constitute de-facto standards for learning goals in the subjects measured, and thus the curricula in different countries have been aligned with the theoretical framework of these assessments.

- In a number of countries the results and rankings of ILSA have become politically high-stakes: Education ministers and governments perceive them as being a test of the success of their policy (e.g., Poland, Germany, and Israel).

As such these international assessments exacerbate the negative consequences that accompany high-stakes tests in different countries ("teaching to the test", diverting resources, etc.).
Benefits of Technology for ILSA

- A computer-based approach to assessments could ensure that tests align better with the performance levels of individual students, giving them more challenging and more manageable and adaptive tasks based on their past performance.

- In particular, adaptive tests allow for a more fine-grained analysis of the entire performance continuum, embracing the most and least able students.

- Use of technology may make assessment more inclusive and fair for all students, including those with disabilities (i.e., by using alternate representations of the same task).

- Finally, computer delivery brings practical advantages, including automated scoring, timely reporting and lower data entry costs.
Variance Between and Within Countries

- Computerized adaptive assessments can provide a more useful profile of countries whose students are clustered in particularly high or low levels of performance.
# PISA and e-PISA Cycle

<table>
<thead>
<tr>
<th>Countries</th>
<th>Reading</th>
<th>Math</th>
<th>Science</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>PISA 2000</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>PISA 2003</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Problem Solving</td>
</tr>
<tr>
<td>PISA 2006</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>CBAS</td>
</tr>
<tr>
<td>PISA 2009</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>PISA 2012</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Problem Solving</td>
</tr>
<tr>
<td>PISA 2015</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>??</td>
</tr>
</tbody>
</table>

**Recent Developments**

- **ERA**
- **EM**
- **Digital?**

- **Financial Literacy**
- **Problem Solving**

*© 2011 # 34*
PISA: Making Use of Electronic Assessments

Not only to Measure Information and Communication Technology (ICT) Literacy Skills

But also to:

- Allow for a wider range of dynamic and interactive tasks
- Explore more efficient ways of carrying out the main tests of student knowledge and skills in reading, mathematics and science
Adding value by:

- Reducing reading and written expression load
- Motivating students for the assessment task
- Linking dynamic contexts with data interpretation
- Enabling student interaction with the media
- Allowing assessment of aspects of science not available in paper-based forms

Only three countries participated in this first digital PISA assessment:

- Iceland, Denmark and Korea
PISA 2009/2012 ERA

- Reading of **electronic texts** is becoming increasingly necessary and prevalent in society.

- PISA recognized this and integrated the reading of electronic texts into the PISA 2009 Reading (ERA) Framework.

- ERA is aimed at assessing the reading literacy of 15-year-olds using electronic texts.

- It was an international option in PISA 2009 (19 countries participated), and it is an option also for PISA 2012.

- ERA is a 40-minute test using existing school IT infrastructure.
Three Major Differences between Print and Electronic Texts

- Electronic reading is more likely to traverse different kinds of texts from different sources.
  - The PISA electronic reading framework sketches a classification of text forms found in the digital medium, and represents this diversity in the ERA instrument with mixed and multiple texts that require readers to integrate information across several sites or pages presenting information in different forms.

- There is a greater onus on the reader to evaluate the text.
  - This is because electronic texts have not typically undergone the scrutiny that is involved in the publication of print-based text. The implication of the mass of information has major implications for readers’ ability to reflect on and evaluate what they read. Readers need to swiftly evaluate the credibility of information; critical thinking therefore gains even more importance in this medium.

- There is a greater onus on the reader to select and construct the text.
  - In print-based texts, the physical status of the printed text encourages the reader to approach the content of the text in a particular sequence. In contrast, electronic texts have navigation tools and features that make possible and indeed even require that the reader create their own reading sequence.
The definition of mathematical literacy in PISA 2012 explicitly calls for the use of mathematical tools, which include technological tools, to make judgments and decisions.

Since PISA items reflect problems that arise in personal, occupational, social and scientific contexts, a calculator is of assistance in many PISA items. Computers and calculators relieve the burden of computation so that individual respondents' attention can be focused on strategies, concepts and structures, rather than on mechanical procedures.

A computer-based assessment will provide the opportunity to extend the integration of technologies - such as statistical tools, geometric construction and visualization utilities, and virtual measuring instruments - into the test items.
The development of competency in problem solving is a central objective within the educational programs of many countries.

Inclusion of problems which cannot be solved without the respondent interacting with the problem situation:

- It will focus on the cognitive skills required to solve unfamiliar problems in life. These cognitive skills may involve modeling or representation of the problem, creative thinking and critical thinking in solving the problem.
- The assessment items be situated in a wide range of real-world contexts and may not have any definite solution.
- The computer-delivered medium will resemble problem solving in real-life situations: variation in the problem situation during the problem solving process, and interaction between the problem solver and the problem.
The plans for PISA 2015 are proceeding on the assumption that computer delivery will be a significant aspect of PISA 2015.

- **All-around solution**: item development, item review, test compilation, test delivery and administration.

- **Ubiquitous**: extensive range of operating systems, over the internet.

- **Innovation**: Exploit the possibilities that arise from the use of new technologies to assess students' knowledge and skills in everyday tasks and challenges, in keeping with PISA's definition of literacy.

- **Adaptability**: solution should also be adaptable to allow for evolution over the PISA cycles e.g., to assess new domains, to cope with new test designs.
Programme for the International Assessment of Adult Competencies (PIAAC)

Aims at developing a strategy to address the supply and demand of adult (16-64) competencies that would:

- Identify and measure differences between individuals and countries in competencies believed to underlie both personal and societal success.

- Assess the impact of these competencies on social and economic outcomes at individual and aggregate levels.

- Gauge the performance of education and training systems in generating required competencies.

- Help to clarify the policy levers that could contribute to enhancing competencies.
Computer and information literacy refers to an individual’s ability to use computers to investigate, create, and communicate in order to participate effectively at home, at school, in the workplace, and in the community.
ICILS 2013

- The assessment of Computer and Information Literacy (CIL) will be **authentic and computer-based**.

- ICILS will incorporate multiple-choice and constructed response items based on realistic stimulus material; software **simulations of generic applications** so that students are required to complete an action in response to an instruction; and authentic tasks that require students to **modify and create information products using “live” computer software applications**.

- ICILS is an **interdisciplinary assessment** and the modules represent a range of focuses, tasks, & contexts related to CIL.
ICILS Broad Research Questions

- What variations exist between countries, and within countries, in student CIL?
- What aspects of students’ use of computers and other ICT’s are related to student achievement in CIL?
- What characteristics of students’ technological backgrounds are related to student achievement in CIL?
- What individual student characteristics are related to student achievement in CIL?
IEA ICILS and OECD PISA

Using computers in international comparative educational projects to:

- investigate students' ability to use computers to manage information and communicate with others
- investigate students' achievement in reading, mathematics science.

ICILS 2013
- Tests students at grade 8 to monitor achievement at a key point in middle school years
- Uses computers to assess students' understanding of computers and students' capacity to work in this medium
- Includes tasks that employ computer software applications for collecting, transforming, creating and sharing cross-disciplinary information
- Collects contextual information from policy makers, schools, teachers and students to help explain variations in student computer and information literacy

PISA 2012
- Tests 15-year-olds to monitor achievement of students approaching the end of compulsory education
- Uses computers to assess students' achievement in the subject domains of reading, mathematics and science
- Integrates computer technology into test items to the extent that they do not contaminate measurement of achievement in selected subject domains
- Collects contextual information from schools and students to help explain variations in student achievement in selected subject domains
Recent Developments
Cognitively Based Assessment of, for, and as Learning

(CBAL)

A Preliminary Theory of Action for Summative and Formative Assessment

Randy Elliot Bennett

Educational Testing Service
of, for and as Learning
“I'm 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 and entrepreneurship and creativity.”
SMARTER Balanced Assessment Consortium (SBAC)

- SBAC will create state-of-the-art adaptive online exams, using “open source” technology.

- The online system will provide accurate assessment information to teachers and others on the progress of all students, including those with disabilities, English language learners and low- and high-performing students.

- The system will include:
  - the required summative exams (offered twice each school year);
  - optional formative, or benchmark, exams
  - a variety of tools, processes and practices that teachers may use in planning and implementing informal, ongoing assessment. This will assist teachers in understanding what students are and are not learning on a daily basis so they can adjust instruction accordingly.
Partnership for the Assessment of Readiness for College and Careers (PARCC)
Partnership for the Assessment of Readiness for College and Careers (PARCC)

Create an assessment system and supporting tools that will help states dramatically increase the number of students who graduate high school ready for college and careers and provide students, parents, teachers and policymakers with the tools they need to help students - from grade three through high school - stay on track to graduate prepared.

PARCC Goals:

1. Build a pathway to college and career readiness
2. Construct assessments that enable cross-state comparisons
3. Create better assessments
4. Make better use of technology in assessments
5. Match investments in testing with investments in teaching
The Winning Consortia

- The consortia - the Partnership for Assessment of Readiness for Colleges and Careers and the smarter Balanced Assessment Consortium - were awarded $170 million and $160 million respectively, to build assessments that evaluate students based on common-core standards by the 2014-15 school year.

"As I travel around the country the number one complaint I hear from teachers is that state bubble tests pressure teachers to teach to a test that doesn't measure what really matters," said Duncan. "Both of these winning applicants are planning to develop assessments that will move us far beyond this and measure real student knowledge and skills."

Robert Linn on CBAL

A New Era of Test-Based Educational Accountability

- The cognitively-based assessment of, for, and as learning (CBAL) system has many of the features that are envisioned in the “smarter balanced” assessment consortium.

- The CBAL summative assessments are not one time events in the Spring, but rather are spread over several occasions throughout the school year. This is a feature that has attracted a good deal of attention in discussions of the consortia plans.

- CBAL also emphasizes the use of formative assessments by teachers and teacher professional development.

- CBAL is based on a theory of action. It is explicitly intended to have positive effects on teacher practice and student learning.

- A program of validity research consistent with the CBAL theory of action would have to give considerable attention to the consequences of test use.

- The developers of the new systems could learn a great deal from the experience and research findings of CBAL.
The Challenges
Policy Challenges

- Do all the above computerized assessments indicate that the educational world is approaching a turning point regarding the incorporation of technology into large-scale assessment?

- Are schools pedagogically, technologically, logistically and socially prepared for this development?

- What are the implications for educators, and for policymakers?

- What will make this mega investment worthwhile?
Yet to be Resolved

Several issues are yet to be resolved before moving LSA and ILSA to a full blown digital and online system, among them:

- Maintaining trend lines between P&P and digital assessments
- Ensuring gender equality
- Narrowing the digital divide
- School system readiness (personnel, hardware, connectivity)
- Infrastructure adaptation to various languages (including right-to-left languages).
Can a CBAL-like Model be Applied to ILSA?

- The ability to integrate continual formative assessments with periodic summative assessments entails complete alignment of the curricula, the content and standards, with the external and internal assessment tasks.

- The solution proposed by CBAL, according to which formative and summative assessment are fully aligned and integrated via technology, may not be easily applicable to ILSA.
Assessment for Learning Cycle

- Assessment
- Standards
- Learning
- Curriculum
- Instruction

Introduction
CET
Large Scale Assessments (LSA)
International Large Scale Assessments (ILSA)
Recent Developments
The Challenges
Can Continual National Accountability System Fit Well with e-ILSA?

Introduction

Large Scale Assessments (LSA)

International Large Scale Assessments (ILSA)

Recent Developments

The Challenges
Alignment of LSA and ILSA

- It is not clear whether such an alignment is possible, or even desired, with regard to the ILSA tasks and national curricula across different education systems in the world.

- Also, ILSA take place on a cycle that is well suited for international comparisons (3, 4, or 5 years).

- As ILSA results have often become a national goal -- this cycle causes certain irregularity in the organization of the learning and the national large scale assessments (LSA) in years which coincide with ILSA.

Thus, there is a need to better synchronize the national and international assessments in countries where both systems exist.
Concluding Remarks

- It is rather clear that technology will continue to advance and improve both LSA and ILSA in an evolutionary manner.

- Technology even has the potential to revolutionize LSA and its alignment with learning and teaching as proposed by the CBAL model.

However, the extent to which ILSA -- based on a common framework which is often not fully aligned with the national curricula -- can join this revolution remains a pedagogical and strategic challenge.
Thanks!

Q&A?