Listening. Learning. Leading.

Innovative Assessment Systems: The Role of New Technology

Randy Bennett
ETS
rbennett@ets.org

Presentation at the International Computer Assisted Assessment Conference, Southampton, England, July 2011

Overview

- Background
- A meta-proposition and 11 related propositions for using technology in innovative assessment systems
- Some significant issues
- Summary

e-Assessment in US Schools

- Local Level
  - 2010 NCES (*Educational Technology in US Public Schools, Fall 2008: First Look*)
    - 72% of public (state) schools reported providing online student assessment
  - NWEA MAP
    - CAT series administered by >3,000 districts throughout the US
    - Has reading, math, science, and language-use measures
    - Employed as early as kindergarten up through end of secondary school

e-Assessment in US Schools

- State Level
  - Oregon Assessment of Knowledge and Skills
    - Over 1.5 million tests administered online per year in primary and secondary school, including for federal accountability purposes
    - Continuously available CAT series in reading, math, science, and social studies
    - May be taken up to 3x per year
    - Paper has become an exception
      - 650,000 students took the online math test in 2008-2009
      - 461 students took the paper version

e-Assessment in US Schools

- State Level
  - Virginia Standards of Learning Assessment Program
    - Over 30 online assessment titles from primary through secondary school (including for federal accountability purposes)
    - Most tests administered online (2009-2010)
      - 78% (~2.1 million tests) online
      - 22% (~600,000 tests) paper

e-Assessment in US Schools

- State “Race to the Top Assessment Program” Consortia
  - SMARTER Balanced Assessment Consortium (SBAC)
    - 31 states
  - Partnership for the Assessment of College and Career Readiness (PARCC)
    - 25 states
- Each consortium will produce an online test series by the 2014-2015 school year to include both selected response and performance tasks for use in primary and secondary education

e-Assessment in US Schools

- National Level
  - National Assessment of Educational Progress (NAEP)
    - Administered Interactive Computer Tasks in science alongside the 2009 science assessment
    - Administered online writing assessment in 2011
    - Will administer online Technology and Engineering Literacy Assessment in 2014

e-Assessment in US Schools

- National Level
  - Hardware Infrastructure
    - 2010 NCES (*Educational Technology in US Public Schools, Fall 2008: First Look*)
      - Ratio of students to instructional computers with Internet access (regardless of location in school) = 3.1:1
        - » Less than 35% FoRPL = 3.1 (.06)
        - » 75% or more FoRPL = 3.2 (.14)

e-Assessment in US Schools

- Hardware Infrastructure
  - 2010 NCES (*Teachers' Use of Educational Technology in U.S. Public Schools, 2009: First Look)*
    - Ratio of students to computers in the classroom every day = 5.3:1
      - Less than 35% FoRPL = 5.9 (.23)
      - 75% or more FoRPL = 4.7 (.27)
    - Ratio of students to computers in the classroom every day plus those that can be brought in = 1.7:1
      - Less than 35% FoRPL = 1.5 (.05)
      - 75% or more FoRPL = 1.8 (.11)

e-Assessment Internationally

- International Level
  - PISA
    - Administered its Electronic Reading Assessment (ERA) in 2009 in 17 countries “using local school infrastructure”
  - Australian National Assessment in Information and Communication Technology (ICT) Literacy
    - Administered in 2005 and 2008 to national samples in year 6 and year 10 on NAP laptops
    - To be administered in October 2011

e-assessment in US Post-Secondary and Professional Contexts

- Higher Education Admissions
  - Graduate Record Examinations (GRE) General Test
  - Test of English as a Foreign Language (TOEFL iBT)
  - Graduate Management Admission Test (GMAT)
- Occupational and Professional Licensure and Certification
  - Uniform Certified Public Accountant Examination
  - Architect Registration Examination (ARE)
  - United States Medical Licensing Examination (USMLE)

And in the UK …

- "'On-screen assessment will shortly touch the life of every learner in this country,' [Ken Boston] said."

  Guardian.co.uk, Wednesday 21 April 2004

- "Simon Lebus, chief executive of Cambridge Assessment, said that traditional examinations are likely to disappear within 10 to 15 years, to be replaced by computerised testing."

  Guardian.co.uk, Sunday, 12 July 2009

- "Students sitting GCSE and other high stakes examinations on-screen will become a reality sooner than people think, according to a new report … produced by Northern Ireland's exams awarding body the CCEA - Council for the Curriculum, Examinations and Assessment - and one of its English counterparts, Edexcel."

  MailOnline, Wednesday, Apr 27 2011

Why is This Happening?

- Technology is infusing all aspects of society
  - Entertainment
  - Social interaction
  - Education
  - Work
- Educational assessment will inevitably follow suit
  - To gain cost and speed efficiencies
  - To stay in step with the tools individuals use to learn and work
    - The more that technology becomes infused in learning and employment, the less valid paper tests will become

A Meta-Proposition

- The use of technology in assessment should be substantively driven
  - Neither “cool” nor efficiency gain alone are sufficient justification
    - If we focus on efficiency, we may end up with nothing more than the ability to create existing tests faster, cheaper, and in greater numbers without necessarily making them better

Why Aren't Curriculum Frameworks Enough?

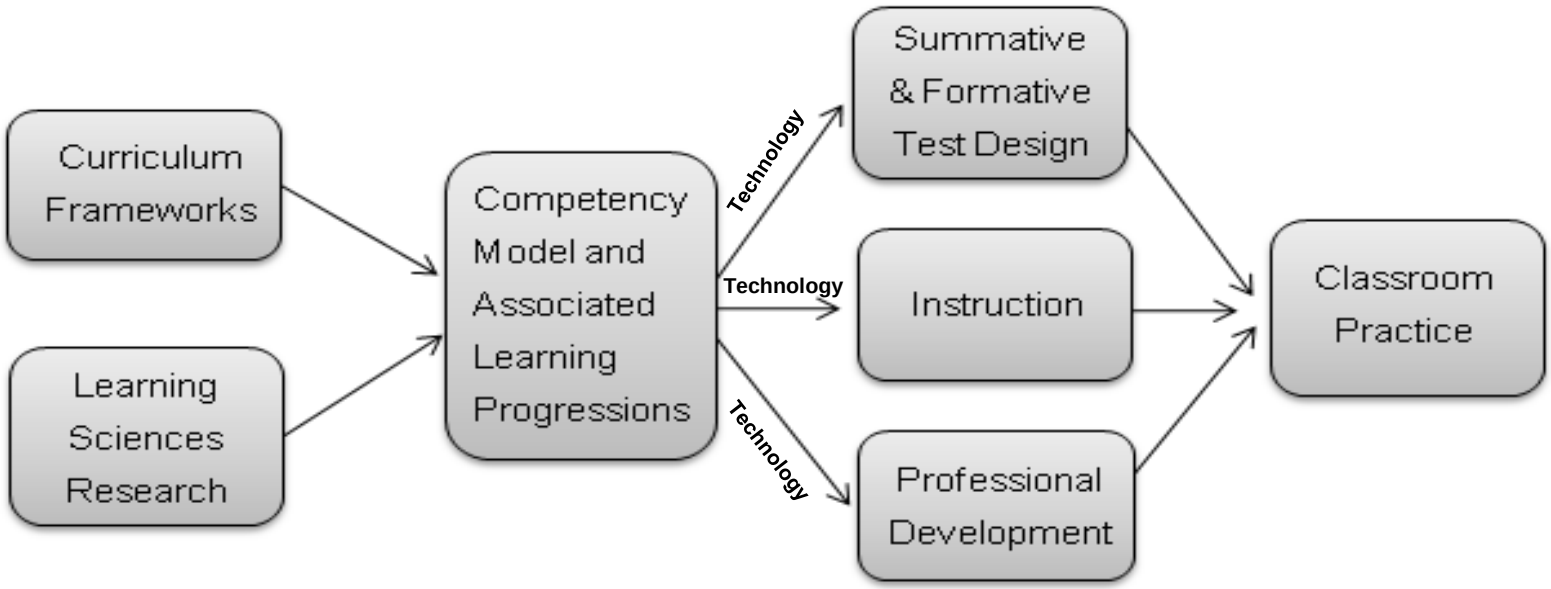
- Curriculum frameworks don't necessarily provide enough guidance to specify how to design:
  - a test
  - (instruction)
  - (professional development)

… in a way that is consistent with the results of learning-sciences research

Substantive Drivers

- The substantive drivers should be a combination of curriculum frameworks with the results of learning-sciences research
  - Principles
    - Describe the characteristics of good learning and teaching practice
  - Competency models
    - Identify the components of domain proficiency (key knowledge, processes, strategies, habits of mind), how they might be organized, and how they might work together to facilitate skilled performance
  - Learning progressions
    - Identify how the components of domain proficiency might be ordered so they can be best addressed instructionally

Driving System Coherence (and Technology Use) Through Competency Models and Learning Progressions



Proposition #1

- Use technology to give students more substantively meaningful assessment tasks than might be feasible through traditional approaches, e.g.,:
  - Tasks that entail significant content for students to reason, read, or write about, or do mathematics with
  - Tasks that students can learn something from

In order to raise money for your annual class trip, you and other members of your class fund-raising committee have decided to sponsor E-Waste Day, a one-day electronics recycling project. People in your community can bring in old cell phones, MP3 players, computers, and other electronic products, which your class will send to a recycling company in exchange for cash. Your committee will need to research the issue of e-waste in order to help the school select the best recycling company for E-Waste Day and learn as much as it can about the issue.

Task 1: Listen to an online radio news report about e-waste and take notes on its content.

Task 2: Evaluate Web sites that contain information about e-waste.

Task 3: Read articles and watch a video about e-waste, and then answer questions about them.

Task 4: Write a paragraph summarizing one of the articles you have read.

Task 5: Examine posters advertising two different e-waste recycling companies. Read paragraphs written by other members of the fund-raising committee, each arguing in favor of using one of these companies for E-Waste Day. Answer questions about the paragraphs.

Task 6: Examine a poster for a third e-waste recycling company. Write a paragraph arguing whether or not to use that company for E-Waste Day.

Task 6A: Help your committee create posters to educate others about the growing problem of e-waste.

Task 7: Improve an announcement about E-Waste Day written by another member of the fund-raising committee.

Task 8: Use a graphic organizer to help manage information from different sources.

Task 9: Using the articles and research notes provided, write a well-informed letter to a computer company.

Proposition #2

- Use technology in assessment to model good instructional practice for teachers, and learning practice for students, by:
  - Including tools and knowledge representations with which proficient performers typically work
  - Encouraging the habits of mind common to proficient performers in the domain
  - Connecting qualitative understanding with formalism

Task: CBAL Reading/Writing Test | Question #: 1 of 1 | Timer: 10 minutes

CBAL

Back | Next

Planning | Chronicle | Gazette | audio | video

Choose a Planning Tool

Click on a work plan below to open it up. When you finish making a plan, click on "Continue to Essay." The planning tool will copy what you have written directly into your essay.

List

Free Writing

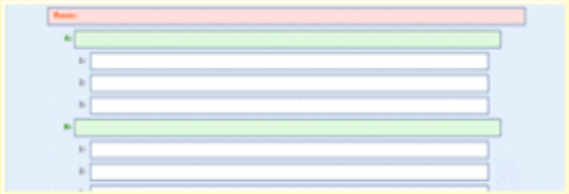
Idea Tree

Idea Web

Outline

Outline

Using an outline is a good way to help organize your ideas about the topic. You can plan your main ideas first and then, under each main idea, you can list some examples, reasons, or details that help support this main idea.



Directions

Task 8: Use a graphic organizer to help manage information from different sources.

Directions: You need to organize your research regarding some of the reasons that e-waste is a major problem in the twenty-first century. Select a graphic organizer to help you organize what you have learned from the articles and media that you have read, viewed, and listened to.

Proposition #2

- Use technology in assessment to model good instructional practice for teachers, and learning practice for students, by:
  - Including tools and knowledge representations with which proficient performers typically work
  - ***Encouraging the habits of mind common to proficient performers in the domain***
  - Connecting qualitative understanding with formalism

Guidelines | Essential Questions

Guidelines for Choosing a Good Internet Source

The information should be clearly **relevant** to the topic.
(If you are not sure whether it is relevant, look again at the "Essential Questions.")

The site should be **trustworthy.**
(If a site is trying to sell you something, the information could be biased.)

The site should be **authoritative.**
(The author of the information offered by the site should be an expert in the subject.)

The information in the site should be **current.**
(An old source might contain information that is no longer accurate, especially when you are researching scientific or political topics.)

Task 2: Evaluate web sites about e-waste.

You have conducted an Internet search and found some sites that might be useful. But are all of these sites worth investigating?

Directions: First, read "Guidelines for Choosing a Good Source." Then go through the List of Sources and decide whether or not each site is likely to give you the information that you need.

1. Choose four sites that you think will NOT be useful by clicking on the DON'T USE box.
2. Choose three sites that you think WILL be useful by clicking on the USE box.
3. Choose two sites that you would need to see more of before deciding whether or not they are useful by clicking on the EXPLORE box.

Click on "Essential Questions" whenever you want to review the main issues.

Use	Don't Use	Explore	Sources
			Does Your PC Hurt the Environment? Computer companies like Dell and Gateway have formal programs for recycling e-waste. More computers have been returned to their manufacturers in 2007 than were returned in 2005 and 2006 combined ...
			Give Us Your Stuff! We pay TOP $ for USED electronic equipment. Great deals on your old electronic gadgets. No product too small! Learn what experts have to say about the benefits of recycling e-waste ...
			Our Future Is Now! Last updated 04-2008. Discussion forum for people interested in serious issues in the environment. Let's talk about making our planet a safe place! If you care about recycling and conservation, this is the forum for you. ...
			China's E-Waste Problem: Confronting the Challenges At the 2008 worldwide conference of scientists in Beijing, participants discussed ways to address e-waste without the problems that come from

Proposition #2

- Use technology in assessment to model good instructional practice for teachers, and learning practice for students, by:
  - Including tools and knowledge representations with which proficient performers typically work
  - Encouraging the habits of mind common to proficient performers in the domain
  - ***Connecting qualitative understanding with formalism***

Dams and Droughts

This picture shows a lake that had a water crisis in 2007.

When the lake is full, you can reach the water from the dock.

The lake has a dam at one end. Water flowing from the lake past the dam is used to create electricity and provide water for crops. The lake may become so shallow that there will not be enough water to generate electricity.

Will the lake become so shallow that water can no longer flow through the dam?

Start with sink: empty

Start with faucet: on full

Start with plug: in

The water will stop when
the first of these settings is reached.

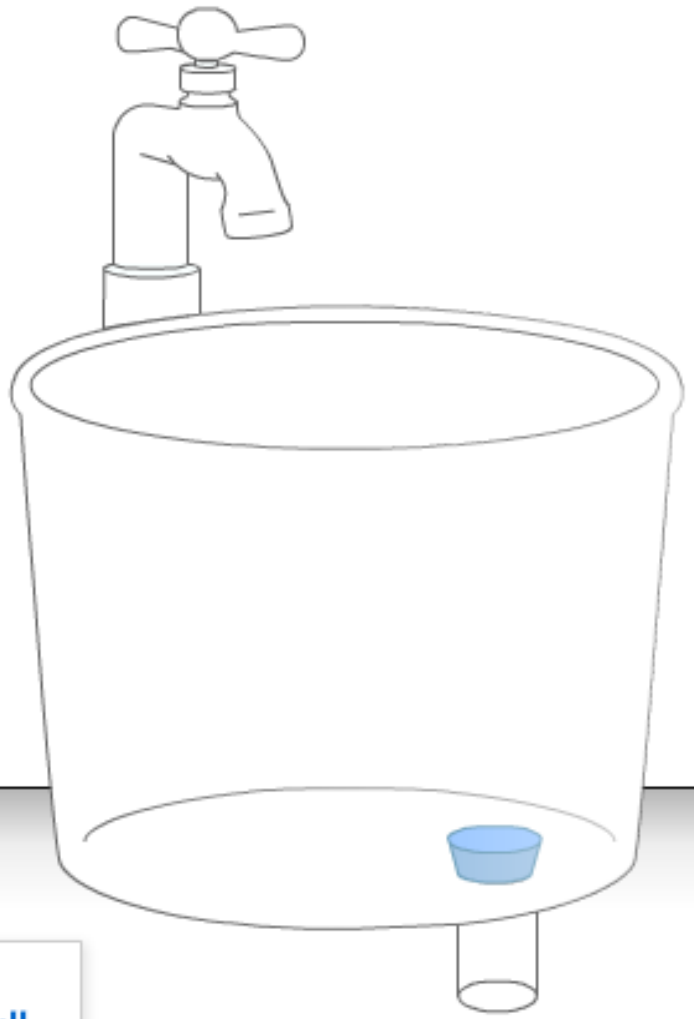
Stop when sink reaches: select

Stop when timer reaches: select

START

RESET

STOP



Timer (min : sec)	Water goes in (gal/min)	Water goes out (gal/min)	Sink contains (gallons)	% Full
0 : 00	3	0	0	0

Proposition #3

- Use technology to assess important (higher- and lower-order) competencies that are *not* measured well in conventional form, e.g., having students:
  - Read orally at the phoneme, word, and continuous-text levels
  - Mathematically model complex problem situations with a spreadsheet
  - Write on computer and read electronic text
  - Use simulations of dynamic systems to interpret evidence, discover relationships, infer causes, and pose solutions
    - Epistemic games

Proposition #4

- Use technology to measure “problem-solving with technology”
  - Successful performance in the workplace and advanced academic settings increasingly requires skill in using technology for cognitive activity

Some scientists study space with large helium gas balloons. These balloons are usually launched from the ground into space but can also be launched from spacecraft near other planets.

Why do scientists use these gas balloons to explore outer space and the atmosphere instead of using satellites, rockets, or other tools? Be sure to explain at least three advantages of using gas balloons.

Base your answer on more than one web page or site. Be sure to write your answer in your own words!

You will be scored on:

- how well you search,
- the quality of your bookmarks, and
- how well you answer the questions

Copy = Ctrl + c
Paste = Ctrl + v
Find = Ctrl + f

NAEP TRE Search Page

Enter your search below:

Tips for searching

balloons Go

If you want to narrow your search, try placing the word "and" between your search terms.

Proposition #5

- Use technology to collect response information that can enlighten substantive interpretation
  - Response latency:
    - Automaticity (e.g., for fluency of procedural skills)
    - Motivation
  - Problem-solving processes

Some scientists study space with large helium gas balloons. These balloons are usually launched from the ground into space but can also be launched from spacecraft near other planets.

Why do scientists use these gas balloons to explore outer space and the atmosphere instead of using satellites, rockets, or other tools? Be sure to explain at least three advantages of using gas balloons.

Base your answer on more than one web page or site. Be sure to write your answer in your own words!

You will be scored on:

- how well you search,
- the quality of your bookmarks, and
- how well you answer the questions

Copy = Ctrl + c
Paste = Ctrl + v
Find = Ctrl + f

NAEP TRE Search Page

Enter your search below:

Tips for searching

balloons Go

If you want to narrow your search, try placing the word "and" between your search terms.

Proposition #6

- Use technology to make assessment fairer for all students, including those with disabilities and English Language Learners, e.g.,
  - Vocabulary links for difficult words (where vocabulary knowledge is not being tested)
  - Alternate representations of the same information (text-to-speech, verbally described graphics)
  - Alternate questions measuring similar skills at similar difficulty levels, when a class of questions is not suitable for some students

Three of the following sets of lengths could be used to construct a triangle. Which set could **not** be used?

(A) 4 cm, 5 cm, 6 cm

(B) 3 cm, 4 cm, 5 cm

(C) 2 cm, 3 cm, 4 cm

(D) 1 cm, 2 cm, 3 cm

Directions:
Use the Control key (Macs use Apple key) to turn the key word feature on and off. When on, click on the green boxes to translate the word.

Three of the following sets of lengths could be used to construct a triangle. Which set could **not** be used?

(A) 4 cm, 5 cm, 6 cm

(B) 3 cm, 4 cm, 5 cm

(C) 2 cm, 3 cm, 4 cm

(D) 1 cm, 2 cm, 3 cm

Translator

English

Español

Directions:
Use the Control key (Macs use Apple key) to turn the key word feature on and off. When on, click on the green boxes to translate the word.

Three of the following sets of lengths could be used to construct a triangle. Which set could **not** be used?

(A) 4 cm, 5 cm, 6 cm

(B) 3 cm, 4 cm, 5 cm

(C) 2 cm, 3 cm, 4 cm

(D) 1 cm, 2 cm, 3 cm

Translator

lengths:

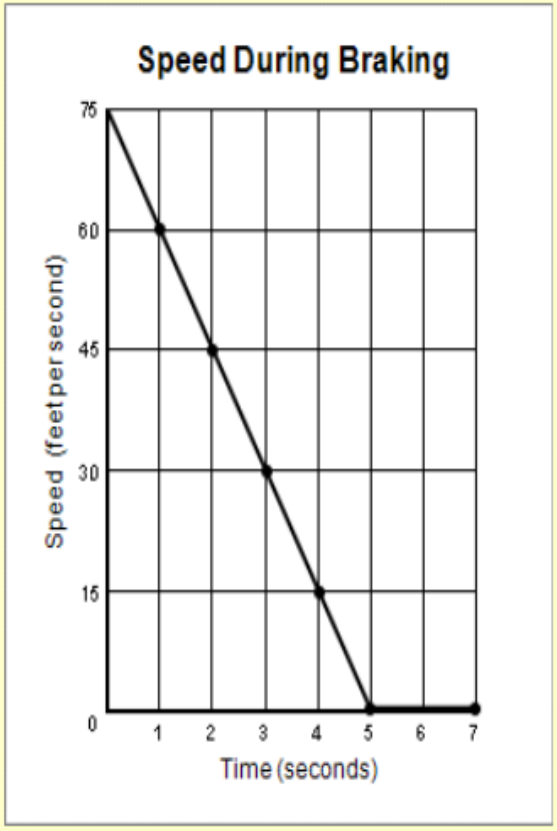
rayas

Directions:
Use the Control key (Macs use Apple key) to turn the key word feature on and off. When on, click on the green boxes to translate the word.

Proposition #7

- Explore the development of adaptive testing approaches more capable of assessing the full range of important competencies than current approaches
  - Adaptive tests adjust their difficulty dynamically to the skill level of the examinee
  - Conventional CAT has been built around MC items because real-time scoring is required to adapt
- Potential Improvements
  - Including automatically scorable (short) CR items
  - A conventional adaptive test that ultimately routes students to an appropriately difficult ECR section (with human scoring done post-administration)

The graph shows the speed of the same car over a time interval on the same dry road after the brakes have been applied. The car's speed is decreasing from 50 miles per hour (about 75 feet per second).



Let

t = time in seconds

s = speed in feet per second

a. Since the line is decreasing, the slope should be negative .

b. What is the value of the slope of the line?

15 feet per second per second

c. Write an equation using t and s that shows how to find the speed, if you are given the time.

$s = -15t + 75$

Proposition #8

- Use technology to measure more frequently, aggregating information over time to form a summative judgment
  - Sensible for large-scale assessment programs only if those more frequent measurements:
    - Contain substantively meaningful tasks
    - Model good instructional and learning practice
    - Provide useful interim-progress information
    - Offer (in the aggregate) better information for decision making than a single end-of-year test

Proposition #9

- Use technology to improve the substantive aspects of scoring
  - Onscreen marking allows real-time monitoring of examiner performance
    - Score characteristics
      - Too lenient, severe, narrow
    - Timing
      - Too quickly
    - Annotation
      - Inappropriate response features

Proposition #9 (con't)

- Use technology to improve the substantive aspects of scoring
  - Automated scoring
    - Significant advances in scoring essays; short text responses; math equations, numerical, and graphical responses; and spoken-language responses
    - For essay scoring, proxy measures are typically used to *predict* the score a human examiner would assign
      - Practicing the proxies may lead to higher machine scores but not necessarily to improved writing skill
    - Disclose how automated scoring works
    - Focus on more fully representing the key competencies themselves (rather than simply predicting the operational behavior of human examiners)

Proposition #10

- Use technology to report assessment results in a timely and instructionally actionable manner
  - When final results can't be reported immediately, consider "phased reporting"
    - Report classroom-level item performance and common errors as soon as possible
    - Report results later that require human CR scoring, extensive analysis, or higher levels of quality control

Proposition #10 (con't)

- Use technology to report in a timely and instructionally actionable manner
  - Structure electronic reports so that they connect performance to curriculum frameworks in a coherent, comprehensible, and useful way
    - Offer tentative placement in a learning progression
    - Suggest (or provide) formative assessment that might be used to confirm the tentative placement
    - Link to sensible next steps given that placement
    - Point to instructional materials with which to pursue those next steps
    - Link to a description of the characteristics of good performance at the next level and exemplars of what that performance looks like

Proposition #11

- Use technology to help teachers (and students) understand the characteristics of good performance by participating in onscreen marking
  - Students
    - Score their own work, and others' anonymous work, as an instructional exercise
      - Using annotation tools
  - Teachers
    - Score their students' work as part of student formative assessment
    - Score as part of the operational assessment program
    - Score as part of a structured, ongoing professional development experience

PERSPECTIVE™

Support | About Us

Educator Home | Understand Scoring

Overview | Learn About Scoring | Anchor Papers | Practice Scoring

Educator Home > Understand Scoring

Understand Scoring

Understand Scoring will help you by providing:

- Information about the Virginia Standards of Learning English Assessment
- Opportunities to study the rubrics used for scoring the direct writing assessments
- Anchor papers that illustrate each score point of each domain
- Practice scoring several sets of papers written by Virginia students
- Verification of your progress by evaluating the accuracy of the scores you assign to a final set of papers (verification set)

Learn About Scoring

Click here to view overview, introduction to and instructional material about the scoring process.

View Anchor Papers

Click here to view Anchor Papers and annotations about the assigned score.

Practice Scoring

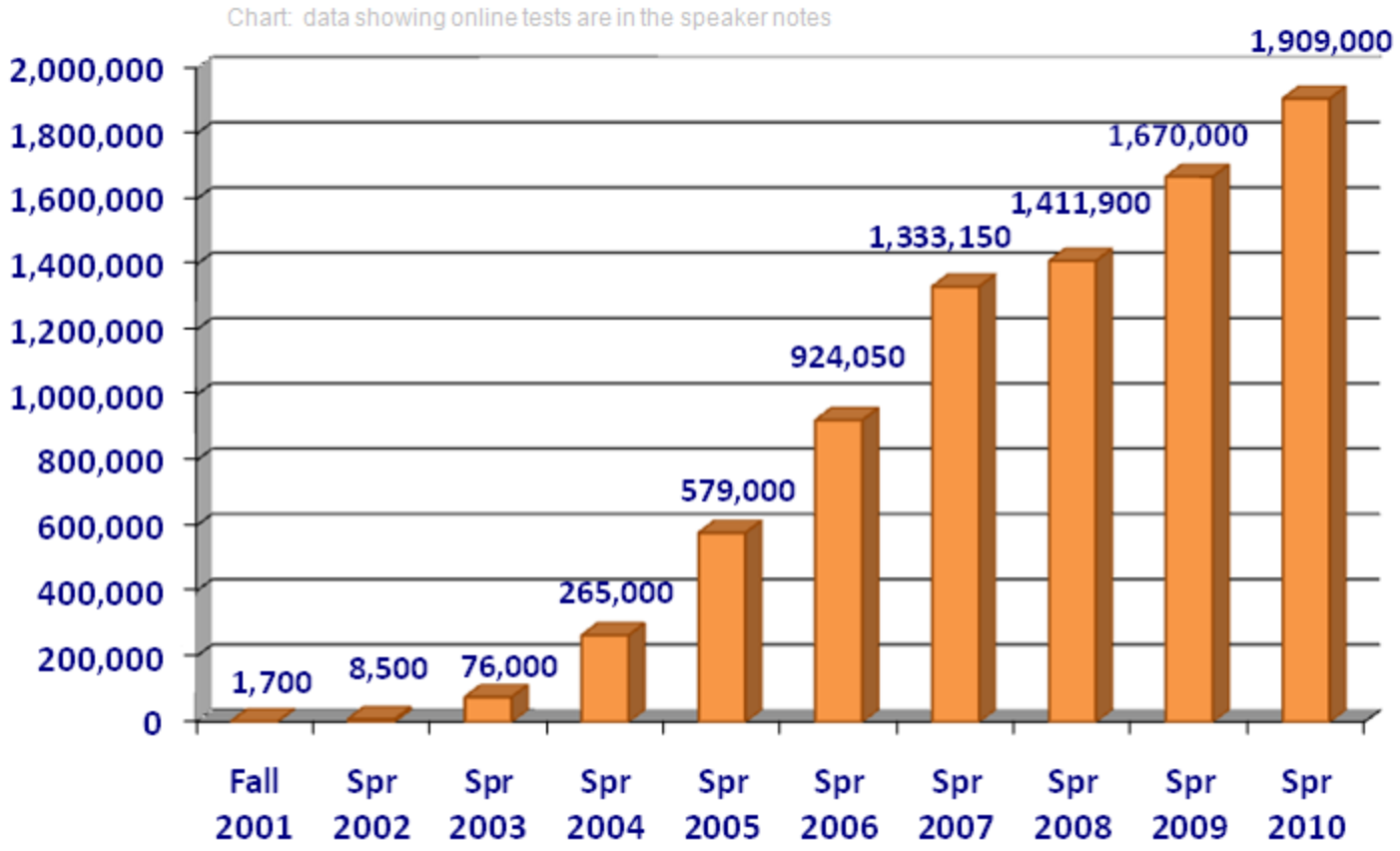
Click here to practice scoring student work and compare your results to assigned scores.

PEARSON

Some (of the many) Issues

- Logistical feasibility
  - National technology infrastructure is not yet adequate to test all students in a cohort efficiently and securely
- Cost
  - Innovative technology-based assessments are very expensive to create
- Fairness
  - The competencies of students with limited computer skills could be underestimated through the use of technology-based ECR tasks
  - Graphics and dynamic stimuli could pose accessibility problems for some students with disabilities
- Scientifically defensible
  - It's not immediately clear how to extract meaning from the additional data that could potentially be collected through highly interactive performance tasks

Online Testing in Virginia: A Phased Approach



2007 – 2008			2008 – 2009			2009 – 2010		
Paper tests:	1,058,623	(39%)	Paper tests:	841,630	(31%)	Paper tests:	595,709	(22%)
Online tests:	1,646,614	(61%)	Online tests:	1,850,013	(69%)	Online tests:	2,104,490	(78%)

Summary

- Technology is being used for primary and secondary school assessment in significant ways at a variety of levels
- Meta Proposition: *The use of technology in assessment should be substantively driven*
- Use technology to:
  1. Give students more substantively meaningful tasks than feasible through traditional approaches
  2. Model good instructional and learning practice
  3. Assess important (higher- and lower-order) competencies not measured well in conventional form
  4. Measure “problem-solving with technology”
  5. Collect response information that can enlighten substantive interpretation
  6. Make assessment fairer for all students

Summary

- Use technology to:
  7. Explore the development of adaptive testing approaches more capable of assessing the full range of important competencies
  8. Support more frequent measurement
  9. Improve the substantive aspects of scoring
  10. Report in a timely and instructionally actionable manner
  11. Help teachers (and students) understand the characteristics of good performance by participating in onscreen marking

Summary

- Significant logistical, cost, fairness, and scientific issues would need to be addressed for these propositions to be effectively implemented on a large scale
  - Limited technology infrastructure
  - Development expense of innovative technology-based assessments
  - Accuracy of assessment for students with limited computer skills
  - Accessibility for some students with disabilities
  - Interpretability of information from highly interactive performance tasks

Listening. Learning. Leading.

Innovative Assessment Systems: The Role of New Technology

Randy Bennett
ETS
rbennett@ets.org

Presentation at the International Computer Assisted Assessment Conference, Southampton, England, July 2011