

Recommendations for Platform Functionality (Question #2)

Randy Bennett
ETS
rbennett@ets.org

Presentation at the Race to the Top Assessment Program
Public & Expert Input Meetings, November 13, 2009, Boston, MA

Question #2

- *We envision the need for a technology platform for assessment development, administration, scoring, and reporting that increases the quality and cost-effectiveness of the assessments. Describe your recommendations for the functionality such a platform could and should offer.*

Recommendation #1a

- The platform should support the development, presentation, and scoring of assessments that represent as fully as possible (1) the standards and (2) the results of cognitive-scientific research (that can help translate the standards to test specifications and to classroom practice)
 - Measurement of higher-order thinking skills (e.g., conceptual understanding, problem solving, reasoning, critical thinking, strategic thinking)
 - Measurement of lower-level components (e.g., declarative knowledge and automaticity)
 - Requires capability to collect timing data
 - Measurement of problem-solving processes (which can be especially useful for formative purposes)
 - Requires capability to collect keystroke and mouse-click data
 - Modeling of the habits of mind characteristic of proficient performers in the domain
 - Requires capability to integrate tools and performance criteria into test questions

Recommendation #1b

- The platform should support the development, presentation, and scoring of assessments that include:
 - Dynamic stimuli (audio, video, animation)
 - Constructed responses (written, spoken, digital representations of artifacts or of performances)
 - Simulations (e.g., of physical or social systems)
 - Information resources (e.g., websites, manuals)
 - Scenario-based, extended exercises calling for the integration of multiple skills and knowledge components
- The platform should also support the development, presentation, and scoring of traditional test questions
- Why?
 - The types of tasks encountered, and competencies required, in workplace and advanced academic settings *cannot* be effectively represented through traditional testing approaches alone

Recommendation #2

- The platform should support frequent measurement, with the capability to aggregate information over time to form a summative judgment
 - Multiple summative tests, distributed across the school year
 - One or more standardized projects
 - Electronic portfolios
- Why?
 - We should be able to make more meaningful (and fairer) decisions about students, teachers, schools, and education systems if we combine evidence from multiple time points and sources

Recommendation #3

- The platform should minimize the influence of irrelevant factors on performance
 - Include tutorials, practice tests, formative assessments, and instructional exercises that use the same interfaces, representations, and tools as found on the summative assessments
 - Account for the needs of students with disabilities and English language learners
- Why?
 - Test performance should depend only upon those aspects of student competency that are the intended targets of measurement

Recommendation #4

- The platform should support (an advanced type of) adaptive testing
 - Traditional item-level adaptive tests require short tasks that are machine-scorable in real time
 - We should look toward new approaches to adaptive testing, e.g.,:
 - A traditional adaptive test section that routes students to an appropriate extended-constructed-response (ECR) section
 - A multi-stage adaptive test in which each stage consists of an extended, scenario-based task including both machine-scorable and ECR items, with routing from one stage to the next based only on the machine-scorable items
- Why?
 - Adaptive tests can provide more precise measurement than traditional tests for low- and high-performing students *but*, in their current form, adaptive tests omit measurement of key competencies

Recommendation #5

- The platform should support online scoring and automated scoring, as well as their combination
 - Online scoring allows for geographically distributed human rating, with real-time monitoring of rater performance
 - Significant advances have been made in the automated scoring of essays; short text responses; math equations, numerical, and graphical responses; and some types of oral responses
 - Detailed performance feedback can often be provided
- Why?
 - These approaches can potentially make scoring cheaper, faster, and better (especially when online and automated scoring are used in combination)

Recommendation #6

- The platform should make it easy to switch testing vendors
 - Represent test questions and automated scoring models in common formats so that questions and scoring models can be moved from one vendor's system to a subsequent vendor's system without undue time, cost, and effort
- Why?
 - States should be able to make vendor selections without having to bear the cost of repeatedly converting test content and scoring

Summary of Recommendations

1. The platform should support the development, presentation, and scoring of assessments that:
 - a. Represent as fully as possible (1) the standards and (2) the results of cognitive-scientific research that can help translate the standards to test specifications (and to classroom practice)
 - b. Include a wide variety of performance-based as well as traditional task types
2. The platform should minimize the influence of irrelevant factors on performance
3. The platform should support frequent measurement, with the capability to aggregate information over time to form a summative judgment
4. The platform should support (an advanced type of) adaptive testing
5. The platform should support online scoring and automated scoring, as well as their combination
6. The platform should make it easy to switch testing vendors