SIIA 2009 Ed Tech Industry Summit – May 3-5, 2009

Harvesting Information from Assessment to Support the Learner and the Educator

Session Description:
Most current uses of assessments fit neatly into two boxes: summative and formative. Technology begins to blur the lines between these types of assessments: not only are assessments possible for a wider range of activities, but it is possible to get timely and relevant feedback, linked to standards and learning activities for every student. What are existing examples of technology-based assessments used to support K-20 students and faculty? Are breakthroughs in technology-based assessments really just around the corner?

Moderator:
Phil Vahey, PhD, Senior Research Scientist, SRI International

Panelists:
Bala Balachander, VP/CTO, CTB/McGraw-Hill
Geneva Haertel, Senior Educational Researcher, Director of Assessment Research Center for Technology in learning, SRI International
Denis Newman, President Empirical Education

Summary by Rabbi Cheryl Weiner, PhD

CURRENT STATE
<u>What are the types of decisions being made using assessment results?</u>
<u>Given these decisions, how is useful information provided to stakeholders?</u>

- The most important aspect of assessment today is that data be collected in the form of "action analytics", information that can be used in a timely manner by stakeholders to make decisions about intervention. People want to know how assessment informs instruction, so that they can provide stakeholders with the right data.
- There is a continuum of assessment from Formative to Summative Testing.
- In formative assessment, there are two types of tests. Prescriptive tests that let stake holders know where they are now and where they should be at a particular time based on that data. Diagnostic tests have benchmark data, which informs instructional decisions about particular individual students or groups. Districts can customize items that can give them specific information about areas that they are interested in.

- In summative assessment, stakeholders are interested in accountability. Aggregated results allow for cross comparisons and can inform key curriculum decisions or instructional shifts across a class, building, or district. The No Child Left Behind (NCLB) program uses this kind of assessment.
- Formative testing is usually sold into districts and Summative testing is sold into states. This indicates the types of decision-making that tests are being used for and by whom. Different statistical analyses are used for different types of tests.
- Stakeholders can use either type of assessment to determine programmatic shifts, as a traffic signal to make decisions about curriculum or goals, red/stop and revise yellow/ need to tweak, green/great.
- Data can be very granular, including information about the distracters (wrong answers) for a multiple-choice item (question). This data can indicate that the item is not clear, that the concept it is testing is not clear, and what students find confusing.
- Score reports need to be qualitative with data reported with actionable interventions. This helps the stakeholders with their decision-making and gives a context to the metrics provided in terms of percentile rankings, etc.

CHALLENGES: (Many of these comments came from the audience, not panelists.)

- Who are the harvesters? Data needs to be provided in different formats, with different interpretations depending on the receiver. Getting data that is actionable, in a timely manner to the constituents who need it---that is what the companies are focusing on.
- Data is often warehoused at the state level without any meaningful impact at the local level. Often the data is aggregated in a way that compares apples to oranges. The state and district have to come to some common ground in determining what they want to know and how they want to use that information to make a difference.
- Organizations can use assessments to make decisions about how schools are performing and how teachers are performing. Many variables other than the teaching/learning relationship can impact the results and have to be taken into account in decision-making.
- Further, any type of “growth model” assessment needs to be carefully examined to make sure that the assumptions are correctly analyzed for decision-making. For example, tests should not be used for personnel decisions in certain situations where teachers share students. This type of assessment is best used for professional development or program assessment. Assessments should not be used punitively.
- The granularity of data that can be given to stakeholders for decisions must be matched by instructional information in order to have an impact on differentiation.
- Legislation can determine who gets stimulus package money based on what forms of assessment are used and how the data is analyzed and this can be a disadvantageous way to distribute funds.
- Tests do not measure higher-level cognitive skills, particularly in math. Inquiry methodology is hard to measure. Portfolio assessment is time consuming and not always effective.

- A disconnect currently exists between where we are as a society in terms of what we need for students to function in the 21st Century and what our tests are measuring.

TECHNOLOGY TRENDS
What are current trends in analyzing and reporting assessment results?
What are current trends in assessment presentation and delivery?

- Trends: Large districts want more formative evaluation techniques for decision-making. Growth scaling for decisions about correlations between programs, teachers and learning activities. Program evaluation to make better decisions about curricula.
- In terms of score reporting, data can be carved up and delivered in almost any way that stakeholders require or request. However, the analytical reports need to be more teacher-friendly. When data are leading to differentiated instruction, teachers need to be given clearer, more explicit direction on how to better use the data instructionally.
- Also, the reporting is not necessarily useful if the classroom unit is still in tact. Using data to move students out of traditional units helps differentiate instruction.
- Reports that stress actionable, practical interventions work best. Presentations of data graphically and with color coding/charts are also easier to understand. Performance progress maps per class with directions for a teacher would be great.
- Statistical analysis. Rausch modeling for data analysis is becoming more prevalent. This and other types of statistical analyses can create better algorithms for determining areas of weaknesses and corresponding interventions for laser like remediation.
- Tests that use speech are being investigated in R & D situations.
- Computer Adapted Tests. These are limited in utility since the platforms to give the tests are not readily available in most schools. Publishers are not rushing to produce these, because stakeholders are not demanding them. Also, the results are not transparent enough to teachers. Since not all children get all items, they don't really understand how to differentiate based on the results. However, these provide much better diagnostic data for lower end students, because you are actually testing at their ability level, the results give better granular information for intervention.
- Cognitive tutoring programs for classroom instruction are useful to teachers/students.
- Ability for teachers to add items/increases value of tests for classroom usage.
- Useful to have the students engaged in their own learning, to embed self-assessment tools in instruction, so students can participate in their own evaluation. Also, to show students test results and ask them about interventions that may work for them. OR perhaps to have a Student Report that is written for them, for them to act on. This lets them see what to do next, how to manage their own instruction.
- Student response devices are prevalent, which should be explored further.

- More use of simulations, animation, and other resources should be examined within the assessment arena and this type of innovation could lead to great stimulus package proposals.
- However, there was a consensus that there is too little innovation using technology. Paper is the predominant model with its inherent lack of ability to be customized on the fly or to allow for scaling. People and policy are not requesting technology.

# # #