

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

Including Pre-Post Assessment in Digital Learning Products

Session Description:

Accountability measures have driven much of the recent product development/sales and marketing for the K-20 education market. Many developers include pre-post assessments to meet the growing need by educators and administrators to manage results. What types of learning products are best served by pre-post assessments? What types of assessments work best in linking assessment and instruction? Will handheld applications impact the way teachers assess their student's learning needs?

Moderator:

Frank Catalano, Principal, Intrinsic Strategy

Panelists:

Bob Ginn, General Manager, Century Consultants, Ltd.

Sue Koch, VP, Marketing, AutoSkill International

Mike Patterson, Senior VP, Assessment Services, Curriculum Advantage

Summary by Sari Follansbee, President, EdTech Design Associates

Including better Assessment in Digital Learning Products

This PowerPoint free session highlighted panelists talking and debating with each other about what works. Accountability measures have driven much of the recent product development, sales and marketing for the K-20 education market. Many developers include pre-post assessments to meet the growing demands of educators and administrators to manage results. But are these a check list item to get a sale or do they really help schools increase teacher and student performance? What types of learning products are best served by pre-post assessments? How do these serve schools? Will handheld applications impact the way teachers assess their student's learning needs?

Frank Catalano introduced the panel and kept the discussion targeted and lively throughout the session. He began asking:

“What are the benefits and risks of adding assessment to curriculum products that didn't previously have it?”

Sue Koch gave a story from AutoSkill, publisher of research-based intervention solutions that help at-risk students build proficiency in the foundation skills of reading and math. They had been clear that they did not do “assessment.” They did pre and post

placement for interventions. What spurred the change was a sales rep coming in (as is so often the case) saying the **Response to Intervention (RTI)** movement was a driving force and AutoSkill had to do something about only being able to enter data manually and deal with in print. This pushed their long-established, research-based business into making their products web based and developing Student Data Integrator, a SIF-compliant software module that seamlessly interoperates with student information systems (SIS).

Sue highlighted key benefits: 1) better, more accurate information about student learning from strong measurable connections between what is tried and how students perform, and 2) more frequent interventions as a result of timely data (the interventions now interconnected with the assessments and reporting).

Risks that Sue highlighted point to difficulties for schools in implementing change. Teachers can be hostile when they don't know what it is – they may need lots of professional development on how to implement. Further, this product training needs to result in increase learning since the data don't lie. Increased access to and understanding of data points to schools' responsibility in providing different approaches and diagnostic tools for different kids.

Mike Patterson stressed the importance of educators being better informed in ongoing instruction – they need rapid low-cost, effective, formative information every day. However, some assessment companies are incapable, slow, or unsuccessful in responding to this need. He underscored that there are too few partnerships between quality assessment and curriculum companies.

In addition to highlighting the benefits of improving practice from marrying assessment with curriculum, Mike spoke about risks in using assessments that are not effective in doing what we expect. For example, if new to the assessment business, test makers may create items that do not measure learning effectively, especially higher cognitive skills. There are several levels of quality; students can get good scores on memorization tests but not really “know” the material, and that may not become clear until they take more substantive benchmark tests.

Mike highlighted the importance of putting the purpose for doing an assessment out front and center. Further he says we need to be honest and recognize that not all assessment material is reliable or comprehensive enough. Psychometric items are expensive and more difficult to create than every-day review items. When assessment materials are not comprehensive and test questions are not field tested, we may or may not be getting a valid picture of student performance.

Bob Ginn spoke about moving from just record keeping to student management. Instead of test item banks that get put in Student Information Systems (SIS), there can be multiple touchpoints – students can be surrounded by their data starting in pre-K. A

key challenge in making this viable is in the transfer of the information – how much data do you maintain and how much/how do you report it back to the teachers?

Bob claims we have done a poor job in giving teachers what they don't already know. We have all the data tools but teachers don't use them – often there is not enough time to administer assessments or to get the data from the individual assessment/curriculum products into the larger SIS systems for either classroom or systemic benefit.

Frank asked, *“What types of assessments are best and how are they developed?”*

Sue Koch noted that AutoSkill partnered to get grade-level passages written and aligned to national norms. They developed their reporting tools in house. Her experience is that customers don't ask how assessments are created, nor can they tell whether a question is valid or not. All they want to know is the student performance results: in this case, words correct per minute.

Mike Patterson reached out and hired people to write the items along with the standards. They insisted on no true/false or trick items, placing emphasis on creating items that can show data over time. What's important and often not done, Mike says, is to pull items that are found to not be valid. What's key is to be clear on your purpose for testing – don't try and make it something that it's not!

Curriculum Advantage didn't want teachers to have to create tests themselves for each week (picking items with standards) so they created the tests ahead of time. They standardized the test taking mechanics and at the same time gave teachers the flexibility to select different sets of items according to their needs. Mike noted that some products where teachers create tests are fine – however, beware that usually this means they can't use data moving forward. Further, quality is affected when teachers aren't taught how to make good test questions.

Bob Ginn spoke about adaptive tests that are much more than a test with x number of items. Since they branch to a multitude of scenarios, record keeping and reporting are difficult. This is an issue when there is demand for reporting out against standards. Bob also highlighted that qualitative test items are hard for any assessment tool because of free text – unless there's the capacity to culling that information down, no one will look at it. He suggested we ask ourselves, “What are we doing manually that we'd like to automate?” We must be clear on what we need to know and the process to do it in order to be able to effectively automate it.

Mike stressed the importance of asking, “What are the purposes for which people are using the test?” because adaptive in and of itself is not necessarily informative. Giving examples from NWEA (Northwest Evaluation Association), he noted that tests must be adaptive against a fixed informed scale – it is not just having the data, but what we can do with that data, that counts.

Regarding formative tests – all agree there is unmet need. It may be undefined but publishers don't want to be left behind.

Frank asked: *“Which data may be transportable?”*

Sue answered that we need to give access to the information that helps teachers teach – that it is important to track the lifetime of each student. This requires the assessments to work with an underlying engine to connect it all – SIS, LMS... – and raises interoperability issues (e.g., SIF certified – only one customer).

Regarding the actual use of SIF, **Bob** noted that at Oracle foundation they got only 3 or 4 requests about SIF – not a big player in this world. **Mike** mentioned that South Carolina at the state level wanted SIF. However, they found it didn't import well, resulting in time delays. State test data from the previous year was uninformative on what the kids can/can't do (knowing whether students are on grade level or not does not tell us what the students know and don't know.) Further, the purchasing paradigm of school districts has not changed. What they want is something economically viable that leverages the teachers' time – the state test is a bi-product not the focus of what they're doing.

Frank asked: *“Clickers, cell phones, other devices as inputs – are they fads, useful, how are they being used?”*

Sue commented that if a device is not connected to other intervention systems then it is not helpful. AutoSkill is not getting demands from customers for clickers, for example, because there's nowhere to connect it to.

Mike said that whole-class responses can give good information. He asks whether being able to see this affects students' ability in taking the test. What is the purpose for using response devices? If simply for feedback, then that may be fine. However, if decisions on students' future are at hand, then these may not be a help, but instead create work that does not go anywhere.

Bob said he found very few teachers were using clickers – it became another thing to have to do. However, like electronic gradebooks, as teachers experience how it is valuable to them, then others will try it.

Sue spoke about problems with voice recognition in dealing with different voices (e.g., southern accent) and therefore it has not become popular for automated marking. Further the need is not high – since there are not enough qualified psychometricians, university professors will moonlight and meet the demand with man power.

Reflecting on assessment in schools, **Mike** noted the paradox of NCLB becoming a byproduct rather than a focus. With so many schools in trouble, they are asking for, and we need to show, alternatives that focus on the students and help them learn.

###