TO: UHH Faculty Congress

FROM: Seri Luangphinit
Chair, Assessment Support Committee

CC: Donald Straney, Chancellor
Kenith Simmons, Interim VCAA
April Komenaka, Accreditation Liaison Officer

RE: Report on the 2011 Conference for the International Association for Educational Assessment (IAEA)

DATE: 18 November 2011

Despite heavy budget constraints and lack of travel funding, the Chair of the Assessment Support Committee felt it important to attend the 2011 IAEA, which is involved with UNESCO, the World Bank, ETS, Cambridge Assessment and other institutions in promoting global standards given the new Common European Framework, oftentimes referred to as the European Qualifications Profile or the Bologna Declaration of 1999 that was signed by 29 countries. According to the British Council, the Bologna Declaration helped further pave the way for the European Credit Transfer System and to impart a stable “European dimension in Quality Assurance comparable criteria and methods.”

The 2011 official release of the Lumina Degree Profile at the 2011 AAC&U Annual Conference is arguably the United States’ answer to this development. This was also the venue for the release of the controversial book, Academically Adrift (2011), which used a limited data pool from the RAND Corporation’s Collegiate Learning Assessment, to present a rather dismal picture of students’ lack of value-added skills upon the completion of tertiary education.

In the months since, the debate has raged over the national call for universal standards and standardized tests to assess those standards. By looking at it from the global context, we can see that this conundrum is not limited to the United States. This international assembly of educational ministries, accrediting bodies, and schools (including primary, secondary and tertiary) provided an ample forum to understand the debate as one not singularly limited to the simplistic “to standardize or not to standardize” debate, but the symbolic role such forms of testing take on a global scale and the reaction

---


of many countries to this current trend towards what is often seen as an increasingly external imposition of bureaucratic management of “their” data. On the other hand, that such systems of management often seem ideal in situations where there is a lack of effective leadership and/or coherence in educational policies on the part of government or governance within institutions.

The following is a brief highlight of key presentations that may be of particular interest to the administration and faculty body at the University of Hawai‘i at Hilo.

**Keynote Speaker:** Robert Sternberg, Oklahoma State University

Dr. Sternberg presented on a number of non-standardized forms of assessments that he has been working on, including Rainbow and Kaleidoscope. He argues that traditional standardized tests only give some information on academic skills but cannot fully predict other life skills that are needed for students to succeed in the real world. In fact, the problems we see in global leadership today stem not from the fact that leaders are unintelligent (many of them graduated from the most prestigious institutions) but that they lack core skills—creativity, invention, facilitation, persuasion, empathy, sensitivity to multiple perspectives, understanding short- and long-term implications of actions, ethical behavior, and a respect for the common good. In many ways, our assessments that singularly target cognitive skills (i.e. critical thinking) neglect these other areas.

He gave the following examples concerning admissions into a graduate program at his home university:

**Student A:** High GRE scores and high BA grades but failed by the end of the 4th year because he while he could recall other people’s ideas and theories, he could not come up with ideas and a project of his own. He also suffered from an immense ego problem that made him unattractive on the job market.

**Student B:** Extremely low GRE scores and average grades but already had a number of publications at the time of application. Student B as not accepted but hired by Sternberg as a research assistant. She was eventually admitted into the PhD program and went on to become one of their most successful graduates in terms of job placement (tenure-track position at a high ranking school) due to her work ethic.

**Student C:** Average GRE scores and average grades but was admitted and instead of becoming an academician, used her training to go to Africa to develop very successful non-profit organizations, suggesting a high level of practical applicability—she could take what she leaned in the classroom and apply it in the real world.

Sternberg then gave examples of “Rainbow Measures” that were funded by the College Board:

- Show students a film of an ethical situation then ask them to rank a series of answers in response to that clip
- Ask students to create a title to a cartoon
- Show students a collage of images and ask them to write a story
• Give students a list of ten “titles” and ask them to write a story around what they pick

Sternberg showed data that when combined with SAT scores, this additional data increased prediction of success by 50%.

When funding was cut for this project, Sternberg undertook the Kaleidoscope project, which developed a series of essay prompts built around measuring practical and wisdom-based skills outlined above.

• Prompt A: What if Rosa Parks had given up her seat on that bus. Can you imagine what history might have been like had this event not occurred? So pick a major historical event and write an alternative history and how such a history might affect you? (Creativity and Understanding short- and long-term implications)
• Prompt B: Describe a moment when you took a risk. What did that risk-taking teach you and what did it help you to accomplish? (Practical application)
• Prompt C: Imagine a project that you could undertake to positively transform your community. What would that project look like and how might you get it started? (Facilitation and Common Good)

Sternberg ended his talk by saying that a computer could not fully evaluate many of these key wisdom-based and life-skills. Online assessment could possibly note grammar and sentence complexity, but it cannot rank creativity or empathy. His argument is for teachers and institutions to develop assessments that can engage these under-valued aspects of a student’s personality, work ethic, and world views.

As a side note, I had the chance to engage in an individual discussion with Sternberg, and he seemed unaware of the events that had transpired at the 2011 AAC&U Conference. He and I both agreed that having Academically Adrift promoted in such a venue without an equally opposing, critical analysis of standardized tests was problematic. As he is a Board Member of the AAC&U, he will be contacting the AAC&U president, Carol Schneider, to request a more inclusive discussion be included in the upcoming Annual Meeting (January 2012) to counter the push being made for the CLA and other standardized assessments.

Assessment in the “Second” and “Third” Worlds

Many countries, ranging from Azerbaijan to Zambia, all reported moving in the direction of direct assessment of student work. The presentations are too numerous to summarize here. However, there was one in particular that best symbolizes the key sentiment among many participants. “Controlling Your Self Control—Myths and Realities of True Assessment in the Global Context” by a panel of two representing the Scottish High International School in New Delhi, whose chair (the main presenter) is also the Executive Director of Hema Engineering Industries. This highly competitive school is at the forefront of educational reforms starting with the National Curricular Framework of 2005 and the Right to Education Act of 2010 that are coming to fruition. The REA stresses the importance of the individual
over the comparison to peers and a holistic approach to student intellectual development as opposed to the limited and limiting cognitive framework (i.e. Bloom’s taxonomy).

This cartoon was part of that presentation and is said to best demonstrate the limitations of standardized tests.

Given this context, it becomes clear why higher level schools like Scottish High International have moved away from standardized tests and now advocate a wider range of assessment techniques, that include: graphic/visual arrangements, interviews (oral interface), direct observation of tasks, portfolios (that include goals set by students; self, peer, parent and teacher assessments; along with samples of what the student considers his/her best work), journals and learning logs.

To stress the necessity for multiple assessment frameworks, they played three songs—a country-western, a classical piece of Brahms, and a pop tune by Shakira. It was made evident to the audience that even a standardized rubric much less a standardized test could not be applied to these three as they are all completely different genres with extremely divergent conventions.

Their move away from standardized tools was also contingent upon faculty realization that entrance examinations were identifying the “best” students, who in turn could not apply their knowledge in real-life tasks. For example, they played clips of classes of students from top levels schools at the fourth grade level who were given the following mathematical question:

Question: The triangle below has a total perimeter of 80 centimeters. The total perimeter of the square is 20. If you take away the square, what is the perimeter of the remaining figure? Explain how you arrived at your answer.

![Triangle and Square Diagram]

A. 60 centimeters  
B. 65 centimeters  
C. 80 centimeters  
D. 90 centimeters

Approximately 85% of students answered A or B, clearly indicating they had confused the concept of “perimeter” with “area.” The majority responded by “logically” explaining that since the square was 20 centimeters, taking it away would mean subtracting the whole or part of that from the overall triangle.
More problematic was when these same children were asked to provide the measurement of a pencil along a ruler. The figure of the pencil was not aligned to the end of the ruler. 65% gave wrong answers—the video clip showed some saying “6,” another said “10” (mistaking the unmarked lines as full centimeters.” So while these children could obviously excel at pre-algebra tests, they still could not properly use a ruler.

Among the many other “myths” they explored involved different types of learning. They asked the audience to assess themselves in terms of learning categories or VAK: kinesthetic (touching and doing), visual (seeing and reading), and auditory (listening and speaking). While most of the audience identified themselves as visual, an actual quiz showed the audience to be otherwise. This suggests that if teachers themselves are not aware of the kinds of learners that they are, they have an even harder time helping students develop their own self awareness of what kinds of techniques work best to develop their learning.

This is why they stressed the need for teachers to be given regular access to professional development and training in up-to-date methods of assessment. Teachers also need to consider the demographics in the classroom. Saini noted that around the world, autism rates range from 1 in 100 to 1 in 35—this has the potential to significantly impact a teacher’s ability to address learning styles on top of already difficult demographic problems such as class, racial and/or ethnic inequalities.

Lastly, Saini took the audience through a rather interesting visual demonstration of what can go wrong in the “traditional” approach to developing an assessment tool. Oftentimes, untrained faculty and administrators are called in to brainstorm; local experts are then called in, often to advocate for standardized or nationally recognized tools and strategies; then sometimes international consultants from first-world nations at immense costs are asked to further develop these tools; then a huge bureaucracy is created to maintain the convoluted system that ends up being more complex and convoluted than what it was ever intended to do—which is to facilitate learning.

The workshop ended with the call for the following considerations when designing effective assessment—utility (is it simple and adequate enough for use?), transparency (is the assessment we do clear to students, parents, and teachers?), “standards” that are matched to what is happening in the classroom and what is already present in the curriculum (negating the need for outside consults to develop lessons and materials), and the development of strategies to move students beyond esoteric exhibitions of Bloom’s taxonomy.
What I Learned from this Conference

I will admit that in speaking with “peers” from South Africa, India, and the Philippines (to name a few), I was immensely humbled by their unwillingness to simply chase after global standards of student success. Even though the IAEA is supposed to highlight the need for more international cooperation and data collection, many participants actually rejected the ideas that were presented by institutions like ETS, Pearson, the World Bank and UNESCO who were calling for more comparative data and centralization for the collection of that data. Interestingly, many African countries were abandoning Cambridge Assessment (the British version of the RAND Corporation) in favor of localized control—this was related to me by Eugenie Rabe, the Chief Operating Officer of the Council for Quality Assurance in General and Further Education and Training, the accrediting body for all primary, secondary and tertiary education in South Africa. Similarly, Dr. Indrani Bhaduri of the Department of Educational Measurement and Evaluation, which is part of the National Council of Educational Research and Training (NCERT) in India, cited their refusal to part with control over any educational data. After all, as she argues, “unless one controls one’s own data, one cannot manage its interpretation, communication, utilization, or verification.”

Such problems were verified by Santosh Gupta, Associate Professor of NCERT. NCERT once contracted ETS to provide training with their IRT model (data analysis) but ETS has never shown the software to the researchers who have legitimate concerns over the way data is collected, analyzed much less disseminated. These concerns lingered even after four members of NCERT—Santosh Gupta, Indrani Badhuri, Veer Pal, and Chandrashekar Konduru—had attended a training session at ETS in Washington.

These individuals have helped me to understand the limitations of internationally much less nationally normed standards, and that even the Lumina Degree Profile needs to be taken with more than a pinch of salt. They also reinforced my belief that assessment should stay firmly in the hands of faculty and not multinational conglomerations. This again begs the need for faculty to receive the training they need—even if it comes from venues like WASC, faculty at some point must be given the right to exercise their own judgment (critical thinking) of what can and will work for them as opposed to having some test or tool imposed onto their teaching.

Nevertheless, Abdullah Al Mamun Bin Quddus of Oxford International School (Bangladesh) related the pressures for individual schools to seek out these internationally recognized educational programs/companies. In the absence of coherent government leadership and in the face of absolutely no collaboration between schools to establish minimum standards of skills, institutions like his are forced to seek external validation through internationally normed measures of assessment because many of their students are looking to compete for admission slots in schools in the United States, the United Kingdom and elsewhere. Similar sentiments were expressed by Janet Evasco of the Center for Educational Measurement, Inc. (Philippines), which administers the National Medical Admission Test (NMAT). Because many students upon graduation from nursing and medical schools go overseas to seek employment, the Philippines is under increasing pressure to guarantee that the qualifications of their graduates meets the same standards of students elsewhere in the world.
I will admit that the presence of multinational test agencies made for a well-rounded discussion. However, one particular presentation by ETS on their Cognitively Based Assessment of, for, and as Learning (CBAL) displayed incredibly precise and rigid scaffolding of Tasks (lessons) that are linked to scaled tests. These tasks revolve around reading, writing, and math, with one particular topic-driven lesson for expository writing that included the following:

- Listen to an online news report
- Evaluate a website
- Read articles on pertinent topic
- Write a paragraph summarizing the materials per written guidelines/rubric on (1) how to do a summary, and (2) how to evaluate a good summary from a bad one
- Examine posters advertising solutions
- Examine a poster from a specific company advertising a solution
- Create a poster yourself
- Create a project that can help your community improve the situation (per the topic)
- Synthesize your ideas into a paper

At this point a member of the Nigerian delegation asked, if you are scaffolding the skills so precisely, then how will students (much less teachers) ever know how to do this for themselves? The ETS representative responded by saying that teachers would have the choice to take away certain steps in the lesson, and that they would be free at any time to customize both the lessons and the tests. The Nigerian delegate thus responded: “But then if students perform poorly on the tests, how would you know if this was due to the customization or the materials themselves? And in any event, you have not answered my question on the problem of the rigid scaffolding you are imposing on the classroom.”

The corporatization of education is well underway in many parts of the world, but in rapidly developing countries like India, whose growth rate exceeded 7% in the last quarter and whose investment in public education has been hovering around 3.1% of its GDP since 2006 (the U.S. is currently spending 5.7% of its GDP), their accomplishments and the refusal of many to buy into the standardized movement is both admirable and humbling (if not downright embarrassing on our part as Americans).