The Report committee for Hemangini Dutt Majumder

Certifies that this is the approved version of the following report

Emergent Practices in the Use of Online Assessment and Measurement to Evaluate Learning

APPROVED BY

SUPERVISING COMMITTEE:

Supervisor: _____

Paul Resta

_____

Min Liu

Emergent Practices in the Use of Online Assessment and Measurement to Evaluate Learning

by

Hemangini Dutt Majumder B.A.; M.B.A

Report

Presented to the Faculty of the Graduate School

of the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

December 2010

Emergent Practices in the Use of Online Assessment and Measurement to Evaluate Learning

by

Hemangini Dutt Majumder, MA

The University of Texas at Austin, 2010

SUPERVISOR: Paul Resta

Abstract:

This report provides an overview of some of the emergent current practices in using technology to evaluate learning. It starts by examining terminology associated with learning evaluation in terms of literature related to the subject. Several innovative models and tools in practice are discussed in terms of their application, situations they are best suited to, advantages or disadvantages they might have and theories they are based on. Some of these are easy to apply and more practically implementable, others are indicative of advanced technologies that are likely to come into use in the future. The report concludes with a few possible scenarios regarding the context in which these technologies and methods are to be used and the real world considerations that would concern the stakeholders.

TABLE OF CONTENTS

TITLE

INTRODUCTION

Purpose of this Report

As the variables educators and learning and development professionals deal with change on a constant basis, new challenges emerge in the field of training and development. Despite the emergence of new learning environments and delivery systems, a continuing theme seems to be the need for measuring learning outcomes and observable value for the investment in learning systems (McCann, 2010). The questions raised through Kirkpatrick's four levels of evaluation i.e. Reaction, Learning, Behavior, and Results (Kirkpatrick & Kirkpatrick, 2006) still hold good. Do people's initial reactions to the learning experience indicate that the learning is relevant to their needs? How effective is the learning and how sustainable will it be? What are people doing differently and better as a result? What results do these investments in learning and development have for the business?

The explosive growth of digital technologies now provides new capabilities, affordances and processes to support and enhance the evaluation of learning and the resulting challenges. The purpose of this report is to explore some of the emergent practices and tools related to online assessment and evaluation and the features that not only enable replication of face-to-face assessments and testing in the online medium but also enrich the experience due to their advanced capabilities.

Definitions and Overview

The words assessment, measurement, and evaluation are often used as synonyms as seen in the works of several scholars. (Carey, 2001; Kubiszyn & Borich, 2003; Popham, 2000; Tanner, 2001; Thorkildsen, 2005; Ward & Murray-Ward, 1999). But when they describe the use of these three concepts, most authors appear to assume measurement is a means to assessment, and assessment is one important component of a much more complicated process called evaluation.

Evaluation involves describing what is and what should be, and comparing the two. To gather information about what is, as well as what should be, assessment is an essential tool. And most assessments involve some kind of measurement process built upon theories of measurement. Scriven (1991), a noted evaluation theorist, summarizes each of these terms:

*"*Measurement [is] a determination of the magnitude of a quantity, typically on a criterion-referenced test scale or on a continuous numerical scale. Whatever is used to do the measurement is called the measurement instrument.*"* Scriven goes on to elaborate on this by describing what an instrument might include (a questionnaire, a test, an eye or a piece of apparatus even an observer). He further explains how measurement fits into the larger scheme of evaluations. "Measurement is a common and sometimes large component of standardized evaluations, but a very small part of its logic, that is, of the justification for the evaluative conclusions. "(p. 266)

"Assessment [is] often used as a synonym for evaluation in which the judgment [usually associated with evaluation] is built into the context of the numerical results." (p. 60). It is significant to note here that the generation of raw scores which have no known content or

construct validity would not be assessment; it is only when the test can generate scores that mathematical analysis can be performed on, that reporting the results constitutes assessment. Terms generally used to refer to evaluation or parts of it include: appraise, analyze, assess, critique, examine, grade, inspect, judge, rate, rank, review, study, test, measure.

As Scriven states "the evaluation process normally involves some identification of relevant standards of merit, worth, or value; some investigation of the performance of evaluands on these standards; and some integration or synthesis of the results to achieve an overall evaluation."

To contrast the concept of evaluation with the measurement process, which also involves the comparison of observations against previously defined standards, we see that measurement is characteristically not concerned with merit, only with 'purely descriptive' properties, and also that those properties are characteristically one-dimensional, which avoids the need for the integrating step. For evaluation on the other hand, the integration process requires some degree of judgment, along with complex calculation,

To summarize, we can conclude that although the terms are often used as synonyms, evaluation uses measurement and assessment to generate value judgments in a wide variety of situations. Assessment usually refers to evaluation of student learning, but it can be used to refer to evaluation in other settings as well. And measurement is usually viewed in support of assessment and/or evaluation.

Similar to the ways new generations of technology have transformed instruction, so have they changed the landscape of measurement, assessment and evaluation. The World Wide Web is becoming increasingly attractive as a mechanism for the delivery of assessment.

"Online assessment in 21st century has come to mean many things to a range of people in various educational contexts" (Northcote, 2002, p. 623). Online assessment is not simply the sum of the Internet and assessment. It also adds significantly to quality of the assessment process in several situations. One view is that online assessments are a better fit for web-based courses than any other methodology. It is also felt that it is better suited for learning targeting 'knowledge' and not higher levels of learning from the Bloom's Taxonomy. Rovai (2000) recommended that computer-assisted testing tools be used in "lower cognitive tasks in low-stakes assessment." These views are largely based on the early history of instructional management systems and web-based courses that have primarily focused on drill and practice and recall learning. However there are several new and more powerful tools that may be used to assess student learning and achievement of higher order skills. Robles and Braathen (2002) stated, "The obvious benefits of online courses may be convenience, flexibility, and 'learning anytime, anywhere'" (p. 39). Harvey and Mogey (1999) also discuss the perceived benefits of online assessments: large numbers can be marked quickly and accurately, student's response can be monitored, assessment can be offered in an open access environment, assessments can be stored and reused, immediate feedback can be given, assessment items can be randomly selected to provide a different set of items to each student. Cassady, Budenz-Anders, Pavlechko, and Mock (2001) found significant differences in performance in the final examination between students who did and did not take advantage of online formative assessment quizzes (p. 6).

According to Bower (2005) "One of the key advantages of online assessment is its capacity to provide retesting opportunities to promote mastery learning" (p. 123). While there are

numerous benefits of implementing technology to manage assessments there are also some limitations that deserve their due attention. Bergstrom, Fryer, and Norris (2005) thought of three main disadvantages of online assessment: Many educators already feel challenged by the task and cost of producing high-end course work delivered with reliable technology and may simply be unequipped to meet the requirements. Providing round-the–clock support to students who are now learning and testing at all hours can be expensive and time consuming. Loss of connectivity can be an issue with online assessment (pp. 62-64).

EMERGENT PRACTICES

This section discusses emerging practices in online assessment. These include the emergence of assessment management systems and issues of interoperability. Computer adaptive testing has been a possibility for a number of years but the emergence of more powerful digital capabilities and tools has led to wider application of this approach to assessment. In addition new tools for online surveys and polls as well as the transcripts of student discourse in discussion forums afford instructors additional means of assessing online learning. Coupled with the growth of online collaborative learning environments, there has been increased interest in the use of peer and self assessment. In addition, there has been expanding development and use of cybercoaching and virtual laboratories in online learning environments. Lastly, the section discusses the emerging field of advanced analytics that enables real time microanalysis of student responses to develop predictive pathways and models of student learning.

Assessment Management Systems and Interoperability Standards

Assessment Management Systems

Assessment Management Systems enable learning and testing professionals to author, schedule, deliver, and report on surveys, quizzes, tests and exams. They included embedded tools to support authoring and categorizing of questions that can be pooled together in a database and reused systematically. There are multiple solutions available in the market that let you store these assessment items locally or in shared or remote repositories (hosted solutions). Assessment management systems like Questionmark enable

instructors to easily create, modify and publish questions and assessments and supports delivery of assessments, management of participants, scheduling assessments, browser-based authoring, and reporting of results (Velan et al, 2002). Respondus similarly supports creating and managing exams that may be used in a variety of elearning systems such as Blackboard, ANGEL, Moodle and other elearning systems. These systems and others allow you to customize your delivery approach in many ways to suit your needs and learning delivery methods. For web based assessments it is possible to deliver assessments to standard or secure web browsers and to mobile devices such as smart phones and PDAs. It is also possible to use these tools to support instructor-led training as they enable printing and scanning of assessments, centralizing management of both online and paper-delivered tests. Assessment delivery options allow questions and choices to be shuffled, and for participants to receive helpful feedback at the item, topic, or assessment level. Options for high-stakes assessment delivery include secure browsers , proctoring security and test center management services. Most solutions also enable administrators to define schedules appropriate for the stakes of each particular assessment, group or individual. When scheduling assessments the instructor can limit dates and number of attempts; require proctor/invigilator log-in; email broadcast schedules to participant and many more options. Other useful features include report generating functionalities that enable the instructor to analyze and share results with stakeholders. Results can usually be exported to multiple formats (Excel, CSV etc) for custom reporting. Most online assessment systems address interoperability by supporting multiple standards (Friesen & Norm, 2005) including IMS, QTI, AICC and SCORM (Descriptions Below).

The IMS Question and Test Interoperability specification (QTI) defines a standard format for the representation of assessment content and results, supporting the exchange of this material between authoring and delivery systems, repositories and other learning management systems (Wills et al, 2009). It allows assessment materials to be authored and delivered on multiple systems interchangeably. It is, therefore, designed to facilitate interoperability between systems. The specification consists of a data model that defines the structure of questions, assessments and results from questions and assessments together with an XML data binding that essentially defines a language for interchanging questions and other assessment material. The XML binding is widely used for exchanging questions between different authoring tools and by publishers. The IMS QTI standards are meant solely for assessment management systems. We will look at some of the other standards in the next few sections.

Aviation Industry Computer-Based Training Committee (AICC)

The Aviation Industry Computer-Based Training Committee (AICC) is an international association of technology-based training professionals. The AICC develops guidelines for the aviation industry in the development, delivery, and evaluation of CBT, WBT, and related training technologies (AICC, 2010). AICC specifications are usually designed to be general purpose (not necessarily Aviation Specific) so that learning technology vendors can spread their costs across multiple markets and thus provide products (needed by the Aviation Industry) at a lower cost. This strategy has resulted in AICC specifications having broad acceptance and relevance to non-aviation and aviation users alike (Walker, 1998).

AICC promotes interoperability standards that software vendors can use across multiple industries. With such standards a vendor can sell their products to a broader market for a lower unit cost. AICC recommendations are fairly general to most types of computer based training and, for this reason, are widely used outside of the aviation training industry.

Sharable Content Object Reference Model (SCORM)

Sharable Content Object Reference Model (SCORM) is a collection of standards and specifications for web-based e-learning that defines communications between client side content and a host system called the run-time environment, which is commonly supported by a learning management system. SCORM is a specification of the Advanced Distributed Learning (ADL) Initiative.

The Department of Defense (DoD) established the Advanced Distributed Learning (ADL) initiative in 1997 to develop a DoD-wide strategy for using learning and information technologies to modernize education and training and to promote cooperation between government, academia and business to develop e-learning standardization. The ADL initiative has defined high-level requirements for learning content, such as content reusability, accessibility, durability and interoperability to leverage existing practices, promote the use of technology-based learning and provide a sound economic basis for investment.

In January of 2000 ADL (ADL, 2010) released the first version of the Shareable Content Object Reference Model (SCORM). And their most recent document that defines sharable learning content objects that meet these high-level requirements is SCORM Version 1.2,

released October 2001.

SCORM features three main sections:

- An XML-based specification for representing course structures, making courses portable between LMS'
- A set of specifications relating to the run-time environment, including an API, content-to-LMS data model, and a content launch specification
- A specification for creating metadata records for content, courses, and raw media elements.

SCORM now also provides a framework to support a framework for assessment (Chang et al, 2004). Several other standards bodies have emerged including: IMS Global Learning Consortium and Electronic Engineers' (IEEE) Learning Technology Standards Committee. ADL and AICC, the two most prominent standards bodies and are now working together (Lifelong Learning Market Report, 2004). AICC is retrofitting much of SCORM's enhancements into its own guidelines and recommendation and SCORM is building on the experience that AICC has gained in three years of certifying LMS product standards. Since ADL does not want to not to compete with the other standards bodies it is attempting to incorporate the work of AICC, IMS, and IEEE under one umbrella.

Certification, compliance and testing remain separate processes. A product that is AICC certified or compliant is not necessarily SCORM compliant and SCORM compliance does not mean AICC certified or compliant. SCORM has been focusing on tracking, tagging, and storing content objects. The standard dwells at length with "metadata" specifying the identifying tags that all learning objects in a course should carry. AICC is interested in

continuing to develop the standards they have already put in place for hardware, operating systems, and audio files. Also, AICC is taking the lead in developing communication protocols or API's to provide a method for one or more programs to communicate with other processes and programs.

The following section looks at the implementation of Questionmark Perception, a leading assessment management tool in a university setting.

Penn State University Implementation of Questionmark Perception

Penn State University, a school located in University Park, Pennsylvania has 19 satellite campuses and a worldwide reach (Questionmark Corporation, 2010). In order to cater to all these students in these remote locations, the school has created and implemented an online curriculum. Dr. Paul Howell, who teaches an online version of his general education introductory class on material science and engineering, not only wanted to make his course available to a wider audience but also to give those on campus more flexibility in how they took the course. Another concern was to safeguard against students who in their eagerness to compete for good grades, might be tempted to cooperate with other students on tests or quizzes.

Using Questionmark Perception the instructor enabled his students to listen to a live lecture or access course information on their computer. The system also enabled all students take the course's weekly quizzes and receive feedback on their computers. Each

week of the semester, students log on to the university's course management system from wherever they are and take the 20-question, end-of-week quiz.

The Penn State faculty and those who manage the database of students and exams take a dual approach to the challenge of keeping tests secure and accurate. The first method, one used by Howell, is to incorporate more openness and respect for the learners into the learning experience. Exams are designed to be open book so the students are allowed to use the course textbook or anything else to solve problems. The emphasis being on understanding the concepts as reflected in the assessment.

Also, there are some other measures enabled by the assessment management system that limit student cooperation. Each learner gets a custom delivered quiz with 20 questions – different from every other student's. These questions are taken from a database of 250 questions for that assessment. The challenge becomes one of ensuring that each student gets questions that are even-handed in both content and level of difficulty.

The school also relies upon Questionmark Perception's ability to save response data across semesters so that they can evaluate each and every question in terms of appropriateness and difficulty. That helped them to reorganize the database in a way that permitted every student to a similar distribution of questions in terms of difficulty level.

Discussion

The first significant benefit we could look at in our discussion of assessment management systems is that they enable learners to possibly get almost instant feedback in some cases

and quicker feedback in terms of scores than traditional assessment techniques (Dermo, 2009).

It is worth considering however, that the most common usage of similar assessment systems is to provide feedback in the form of numbers and scores and, while these do have strengths in terms of quantitative analysis (Ferrao, 2010) and make the results concrete and measurable, it is limited in providing personalized feedback in more qualitative terms to improve the student's performance. There have been attempts to counter this limitation by developing innovative functionalities like looping back to the relevant content in the course if the learner is unable to answer a specific question related to it and others.

As noted earlier, interoperability of learning and assessment management systems also allow data to be shared across different systems and encourage a systems thinking approach. We must remember however, that complete interoperability and seamless integration is still an ideal and there is a long way to go before this can be done as easily as one would like.

Another significant advantage is that assessment management systems allow teachers to focus increasingly on pedagogical content knowledge where significant efforts need to be directed (Dede, 2009). There is less need to get bogged down by repetitive administrative tasks (McLeod, 2009) that take up more time than they should. This frees up significant resources and time to focus on the quality of learning and assessment techniques

There are also certain enhanced functionalities in terms of assessments built to test cognitive skills that have a level of complexity similar to reality. These simulations support 'situated learning' and can potentially provide more powerful ways of assessing more authentic forms of assessment. Despite the potential and opportunities to utilize these tools to address complex and higher level learning needs, it will take changes in current practices to show shift away from multiple choice questions and other knowledge level assessment item types. The templates provided tend to limit the teacher's ability to use these tools to reach their creative potential.

Computer Adaptive Testing

Computerized adaptive testing (CAT) is an innovative test delivery model that, based on student's responses on items, adapts the presentation of subsequent items to the examinee's ability level. CAT successively selects questions so as to maximize the precision of the exam based on what is known about the examinee from previous questions (Weis & Kingsbury, 1984). Compared to static multiple choice exams with a fixed set of items administered to all examinees, computer-adaptive tests require fewer test items to arrive at equally accurate scores. Paper-and-pencil tests are "fixed-item" tests in which all students answer the same questions and give little information about the particular level of ability of each student (Chatzpoulou & Economides, 2010). With recent advancements in measurement theory and the increased availability of microcomputers in schools, the practice of using these tests may change. Computerized tests may replace paper-and-pencil tests in some instances

The basic computer-adaptive testing method is an iterative algorithm with the following steps:

1. The pool of available items is searched for the optimal item, based on the current estimate of the examinee's ability
2. The chosen item is presented to the examinee, who then answers it correctly or incorrectly
3. The ability estimate is updated, based upon all prior answers
4. Steps 1–3 are repeated until a termination criterion is met

The algorithm is generally started by selecting an item of medium, or medium-easy, difficulty as the first item.

There are five essential features of a CAT (the following is adapted from Weiss & Kingsbury, 1984).

1. Calibrated item pool
2. Starting point or entry level
3. Item selection algorithm
4. Scoring procedure
5. Termination criterion

Calibrated Item Pool

A pool of items must be available for the CAT to choose from. The pool must be calibrated with a psychometric model, which is used as a basis for the remaining four components. Typically, item response theory is employed as the psychometric model (Weiss & Kingsbury,1984).

Starting Point

In CAT, items are selected based on the examinee's performance up to a given point in the test. If some previous information regarding the examinee is known, it can be used,[1] but often the CAT just assumes that the examinee is of average ability - hence the first item is of medium difficulty.

Item Selection Algorithm

As mentioned previously, item response theory places examinees and items on the same metric. Therefore, if the CAT has an estimate of examinee ability, it is able to select an item that is most appropriate for that estimate (Wainer & Mislevy, 2000). Technically, this is done by selecting the item with the greatest information at that point.

Scoring Procedure

After an item is administered, the CAT updates its estimate of the examinee's ability level. If the examinee answered the item correctly, the CAT will likely estimate their

ability to be somewhat higher, and vice versa. This is done by using the item response function from item response theory to obtain a likelihood function of the examinee's ability.

Termination Criterion

The CAT algorithm is designed to repeatedly administer items and update the estimate of examinee ability. This will continue until the item pool is exhausted unless a termination criterion is incorporated into the CAT. Termination criteria exist for different purposes of the test, such as if the test is designed only to determine if the examinee should "Pass" or "Fail" the test, rather than obtaining a precise estimate of their ability (Lin & Spray, 2000).

GRE as an example of CAT

The Graduate Record Examination (GRE) is a commercially run (by ETS or Educational Testing Service) standardized test that is an admission requirement for many graduate schools in the United States and in other English-speaking countries. Areas tested include verbal reasoning, quantitative reasoning, critical thinking, and analytical writing skills that have been acquired over a long period of time and that are not related to any specific field of study. In the United States, Canada, and many other countries, the GRE General Test is offered as a computer-based exam administered by select qualified testing centers; however, paper-based exams are offered in areas of the world where computer-based testing is not available.

The Verbal and Quantitative multiple-choice portions of the exam currently use computer-adaptive testing (CAT) methods that automatically change the difficulty of questions as the test taker proceeds with the exam, depending on the number of correct or incorrect answers that are given. The test taker is not allowed to go back and change the answers to previous questions, and some type of answer must be given before the next question is presented.

The first question that is given in a multiple-choice section is picked to be a medium difficulty question that half of the GRE test takers should answer correctly. If the question is answered correctly, then subsequent questions become more difficult. If the question is answered incorrectly, then subsequent questions become easier, until a question is answered correctly.

Ideal Computer Adaptive Tests are of variable length, where the test will stop itself once it has identified a candidate's ability level. However, this effect is moderated with the GRE because it has a fixed length. The actual scoring of the test is done with item response theory (IRT).

Note: In December 2009, ETS announced plans to move forward with significant revisions to the GRE in 2011 (Educational Testing Services, 2010). Changes include a new 130-170 scoring scale, the elimination of certain question types such as antonyms and analogies, the addition of an online calculator, and the elimination of the CAT format of question-by-question adjustment.

Discussion

In general, computerized testing greatly increases the flexibility of test management. Test administrator differences are eliminated (Thissen & Mislevy, 2000) as a factor in measurement error. The test taker is continuously faced with a realistic challenge--items are not too difficult or too easy. Students' performance over time can be tracked by using the computer to store performance data.

In addition to having the advantages of computerized testing, CATs increase efficiency. Significantly less time is needed to administer CATs than a fixed-item test since fewer items are needed to achieve acceptable accuracy. CATs can reduce testing time by more than 50% while maintaining the same level of reliability. Shorter testing times also reduce fatigue, which can be a significant factor in students' test results. CAT technology allows test takers to receive immediate feedback on their performance.

On the flip side, hardware limitations may restrict the types of items that can be administered by computer. Items involving detailed art work and graphs or extensive reading passages, for example, are hard to present using the types of computers found in most schools.

Review of past items is generally disallowed. Adaptive tests tend to administer easier items after a person answers incorrectly. Supposedly, an astute test-taker could use such clues to detect incorrect answers and correct them. Or, test-takers could be coached to deliberately pick wrong answers, leading to an increasingly easier test.

Calibration of the item pool can be a problem. In order to model the characteristics of the items (e.g., to pick the optimal item), all the items of the test must be pre-administered to a sizable sample and then analyzed. To achieve this, new items must be mixed into the operational items of an exam (the responses are recorded but do not contribute to the test-takers' scores), called "pilot testing," "pre-testing," or "seeding (Thissen & Mislevy, 2000). This presents logistical, ethical, and security issues

Online Surveys, Polls and Discussion Forums

There are several online survey tools that are available and can help facilitate the student feedback collection process. These have multiple question types, existing templates for beginner survey creators, customization and data analysis features.

Online surveys, polls and ballots give learners the opportunity to comment on aspects of e-learning design and delivery quickly and have the best response rates (Kirkpatrick & Kirkpatrick, 2006). In live virtual classroom sessions you can use the built in polling feature to ask for immediate feedback on the quality of presentation and delivery. Online testing and survey tools can also be used to post ballots and these ballots can then record scores over a period of time. The following is an example of the use of a survey tool in a university context.

Survey Monkey used at Northern Illinois University

This was a survey administered to a class of students (Looney, 2008) during the third or fourth week of the semester at the department of Kinesiology and Physical Education at Northern Illinois University to help the instructor judge if her pace and clarity in presenting material was appropriate and also whether she had been targeting the students' ability level correctly. This was a major concern for general education classes where the class profile ranged from freshmen to graduate students from all majors. The open-ended items helped the instructor determine what was working well and the areas needing improvement.

The author chose to use SurveyMonkey.com (http://www.SurveyMonkey.com) because of a colleague's recommendation, its ease of use, and the free of charge status for a basic membership.

Survey monkey compiled summary reports for each item, which saved time and allowed easy access to student feedback. After reflecting upon the results, the instructor presented the closed-ended summary results to the students and reported general themes from the open-ended comments. At this time, the instructor addressed any changes that were going to be made or reasons why some suggestions would not be implemented. The students' reactions were positive regarding the process, and their feedback brought about positive changes. For example, students indicated that they had trouble finding homework problems and other items on BlackboardTM (http://blackboard.com). As a result, the instructor reorganized parts of BlackboardTM immediately.

The instructor used the student feedback to make warranted adjustments to teaching and

also presented the information in her annual report to the department's Personnel Committee.

Online Discussion Forums

An important tool for students to share their ideas and to be challenged with multiple perspectives is to engage in online discussion forums (Kirkpatrick & Kirkpatrick, 2006). Such forums may be in the form of a bulletin board where instructors can post questions or issues for learners to respond to. In such discussions, learners can see other learners' comments and respond to them, creating ongoing discourse focused on topics or issues related to the course. Discussion forums are a common feature within learning management systems and online meeting tools and are also available as standalone online discussion tools. You may also use social networking tools such as wikis and blogs to encourage dialogue and discourse. The instructor can view the online discussions to determine the perspectives of the students and course concepts they are having problems with (Mazzolini & Maddison, 2007). To make the online discussions productive, it is helpful for the instructor to seed the discussion with questions that provoke meaningful discussion.

Focus groups that traditionally require a lot of travel and setup time can be replicated quite easily through discussion forums or by the use of synchronous text-based communication called a chat. Chat also has the advantage of leaving behind a written record and there are no notes to transcribe.

Discussion

There are many skeptics of surveys and direct feedback from learners (Kirkpatrick & Kirkpatrick, 2006). It is often opined that these are not accurate representations of whether the learning experience has resulted in actual learning (Duda & Nobile, 2010). It should be considered however that e-learning is often a new experience for learners. These methods (pertaining to Level 1 of Kirkpatrick's model of evaluation) can help to assess any glaring holes in the design of learning and also gauge emotional acceptance of the new initiative. The testimonials and statistics can be important to generate a positive buzz around e-learning.

Another interesting aspect to consider is confidentiality and anonymity. In order to receive honest feedback from learners it is important for them to feel confident that their interests will not be affected in any way by their answers. In traditional paper based surveys one can hide one's identity by not filling out personal information. In most online surveys (Looney, 2008) it is possible to alter settings so that your student does not have to reveal her identity but in the learner's perception there is scope for tracking the response providers identity through email addresses where you send the survey and IP addresses. This might affect the feedback they provide unless trust is established through prior dialogue and would defeat the purpose of the exercise.

The feedback you receive from a poll will differ significantly from the feedback you receive from an online survey and that in turn differs significantly from an online discussion forum or an online chat focus group. It is important to choose your method based on the kind of detail you require and whether you would require this information to

perform further quantitative or qualitative analysis for decision making purposes or to generate ideas and suggestions.

It is important to focus on your learner when designing your survey. It is best to keep the language used direct and use piping and skip logic features to ensure that learners do not have answer long lists of questions that they have already stated are irrelevant to them. Since, there is really no incentive for the learners to provide good feedback (unlike in test scores where there is a personal display of mastery involved) one would do well to make the learner feel that some thought has gone into their convenience and attention span (Singh et al, 2009).

Peer and Self Assessment

Peer Assessment

Since there is great emphasis on working in teams in most classrooms peer assessment has emerged as a best practice. However, if they are to offer helpful feedback, students must have a clear understanding of what they are to look for in their peers' work (Sluijsmans & Moerkerke 1999). The instructor must convey expectations clearly to them before they begin.

The instructor may provide a sample writing or speaking assignment. As a group, students determine what should be assessed and how criteria for successful completion of the communication task should be defined (Searby & Ewers, 1997). Then the instructor

gives students a sample completed assignment. Students assess this using the criteria they have developed, and determine how to convey feedback clearly to the fictitious student.

Students can also benefit from using rubrics or checklists to guide their assessments. At first these can be provided by the instructor; once the students have more experience, they can develop them themselves (Brindley & Scoffield, 1998). It helps the peer evaluator focus on these areas by asking questions about specific points, such as the presence of examples to support the ideas discussed.

For peer evaluation to work effectively, the learning environment in the classroom must be supportive. Students must feel comfortable and trust one another in order to provide honest and constructive feedback.

Self Assessment

In this kind of reflection, students step back from the learning process to think about their learning strategies (Resta et al., 2002) and their progress as learners. Such self-assessment encourages students to become independent learners and can increase their motivation.

The successful use of student self-assessment may be examined through the following:

1. Goal setting
2. Guided practice with assessment tools

3. Portfolios for self assessment

Goal setting

Goal setting is essential because students can evaluate their progress more clearly when they have targets against which to measure their performance. In addition, students' motivation to learn increases when they have self-defined, and therefore relevant, learning goals (Morisano & Shore, 2010).

One way to begin the process of introducing students to self-assessment is to create student-teacher contracts (Bartlett, 2004). Contracts are written agreements between students and instructors, which commonly involve determining the number and type of assignments that are required for particular grades. For example, a student may agree to work toward the grade of "B" by completing a specific number of assignments at a level of quality described by the instructor. Contracts can serve as a good way of helping students to begin to consider establishing goals for themselves as language learners.

Guided practice with assessment tools

Students do not learn to monitor or assess their learning on their own; they need to be taught strategies for self monitoring and self assessment. The instructor models the technique (use of a checklist or rubric, for example); students then try the technique themselves; finally, students discuss whether and how well the technique worked and

what to do differently next time. In addition to checklists and rubrics for specific tasks, students can also use broader self-assessment tools to reflect on topics they have studied, skills they have learned, their study habits, and their sense of their overall strengths and weaknesses.

Students can share their self-assessments with a peer or in a small group, with instructions that they compare their impressions with other criteria such as test scores, teacher evaluations, and peers' opinions. This kind of practice helps students to be aware of their learning. It also informs the teacher about students' thoughts on their progress, and gives the teacher feedback about course content and instruction.

Portfolios for self-assessment

Portfolios are purposeful, organized, systematic collections of student work that tell the story of a student's efforts, progress, and achievement in specific areas. This purposeful collection of work includes the learner's selection of content, the criteria for determining merit and the student's self-reflection (Torrance, 1997). One of the most important dimensions of portfolio assessment is that it should actively involve the students in the process of assessment (Tierney, Carter, & Desai, 1991). The student participates in the selection of portfolio content, the development of guidelines for selection, and the definition of criteria for judging merit. Portfolio assessment is a joint process for instructor and student. There are a number of both commercial and locally developed e-portfolio systems that are used in higher education (McFarlane, 2003). In addition e-

porfolio capabilities are also included in other commercial systems designed to address a broad range of higher education needs.

eVIVA by Ultralab

The eVIVA project (Walton, 2005) was developed at Ultralab in the United Kingdom with an intention to create a more flexible way of assessment, taking advantage of the possibilities of mobile phone technologies and web based formative assessment tools. Ultralab used these tools to promote self and peer assessments as well as dialogue between teachers and students. Students had access to the eVIVA website where they could set up an individual profile of system preferences and record an introductory sound file on their mobile or land phone. Students then carried out a self assessment activity by selecting a series of simple "I Can" statements designed to start them thinking about their abilities. The website also had a question bank which the students chose questions from for their telephone viva at the end which was scheduled by them. They also had their own e-portfolio space in which they were asked to record significant milestone moments of learning and to upload supporting files. Teachers and students would later discuss these at length. Finally the students would dial into eVIVA and record the answers to questions they had picked out before. The teachers made a holistic assessment of the students capabilities based on the variety of measures discussed earlier.

Discussion

A significant benefit of peer and self-evaluations is that students can develop lifelong

evaluation skills both about their own work and thinking as well as others. This allows them to take their first steps towards independent and autonomous learning by developing learning strategies based on their evaluations (Dochy et al., 1999; Mowl & Pain, 1995; Topping et al., 2000). This marks a distinct shift from the behaviorist tradition where the learner is seen as a passive receiver of knowledge delivered.

Keeping the students focused on improvement rather than judgment alone is crucial for personal development. Students need constructive feedback to help them assess and then build on their own strengths. They need to identify ways of addressing weaknesses and plan appropriate action. The key benefit of peer assessment is the opportunity it provides for students to provide each other with multiple perspectives and lateral constructive suggestions (Heron 1988, p. 86). Student criticisms of self and peer assessment include the fear of personal bias - the approach is seen to lack objectivity. It is important to help students see the value and validity of the approach to assessment ('face validity').

The validity of student' relative to teacher' judgments in assessment has been explored by many researchers and reported in the education literature. In large part student peer and self grading works best when students share a comprehensive understanding of the assessment criteria and the characteristics of work illustrative of different levels of performance. They also need to know that you are seriously seeking their constructive input as assessors.

Cybercoaching

The cybercoaching model emphasizes the process of development rather than a procedure for evaluation. Personalized feedback is provided by the instructor by using easily accessible technologies of electronic mail and word processing. The feedback is for the purpose of improving student performance and self-regulation. The technology used is powerful and simple in its approach, not requiring very advanced user skills and that makes this model very practicable. Word processing programs are now almost universally readable, even by competing operating systems (i.e., Mac vs. PC).

Coaching is defined here to include a tutorial function of monitoring student progress to provide feedback for improvement (Clark, 2004). A coach, unlike a referee or a spectator or even a player, looks for skills to develop rather than errors committed or goals scored. It is interesting that the "cyber" part of cybercoaching refers not only to online technology or cyberspace through which the coaching occurs, but is also a reference to field of cybernetics. The field of cybernetics (Smith & Smith, 1965) points us towards the theory of individual and organizational learning based on the need for feedback to adjust the momentum and direction of progress.

A more continuous and cyclical view of the teaching and learning process is advised (Chang & Peterson, 2006). A process of continuously monitoring and adjusting effective instruction would be beneficial. This is the difference between formative and summative assessment. Formative assessment is the coaching feedback provided before the summative, or final, evaluation. The assessment of student learning is also used by the instructor to self-assess the effectiveness of all the instructional decisions. The teacher then returns to the teaching

activity because feedback is an opportunity to extend and personalize instruction.

As with all instructional designs, cybercoaching requires thoughtful organization of teaching activities and learning activities.

Teaching Activities

Teaching activities are the specific ways students will get new information directly or indirectly. Within the cybercoaching model, this includes feedback inserted into the students' work. When the objective of the course is to develop more advanced thinking than simple memory or comprehension, or when the student lacks confidence in making creative decisions, feedback is certainly warranted (Fluckiger et al, 2010). The feedback may direct the student to instructional materials or elaborate on previous instruction. Feedback provides additional scaffolding for students' continued learning.

Learning Activities

As mentioned above, effective instruction includes interaction between instructor and student, and among students. Specific to cybercoaching, the learning activities feature the submission of early drafts or completed work for feedback. The assigned learning activities may include any step of the development process: to develop projects, to submit drafts, to respond to feedback, to revise drafts, or to reflect on the process and formally self-assess progress according to an agreed-upon set of criteria (Dochy et al, 1999). The short-term achievement required for success in the class is not as important as the long-term usefulness. This requires independent decisions by the student to use that learning.

Self-Regulated Learning

The cybercoaching model addresses a core belief that students must develop skills of analysis and dispositions of self-improvement. There is a strong emphasis on metacognition (Bandura, 1997), that is introspection and assessing one's own performance, and analyzing it according to criteria defined in rubrics.

To summarize, these are the salient components of successful cybercoaching:

- Foster a coaching relationship with each student.
- Design rubrics that describe novice to expert performance of each criterion of success.
- Establish a routine of immediacy for feedback, and respond to feedback from students.
- Encourage revision for mastery, separating feedback from grades.
- Respect the limitations of both the word processing and e-mail software. Prepare for an increasingly paperless culture.

Cybercoaching implemented by NEATE (Marra, 2004)

New England Association of Teachers of English (NEATE) is an affiliate of the National Council of Teachers of English (NCTE) that believes in fostering a professional academic community for New England teachers. As a service to new English language arts teachers and veteran teachers with new assignments or interests - NEATE offers cyber coaching on a range of topics by NEATE members who are experienced ELA professionals. Their website links to a database of potential cybercoaches that the learner seeking help can choose from based on their expertise or background.

Discussion

As we discussed in the introduction to this report, it is often felt that while computer-assisted assessment—that is, the use of computers to deliver, mark, and analyze assignments or examinations—has a place (Rudner & Gagne, 2001), they may only be used when there are simple solutions, or 'one right answer' (Stiggins, 2005). This suggests that computer-based tests, however helpful, are limited to testing lower levels of thinking (see Starko et al., 2003). Cybercoaching is very well suited for personalized feedback appropriate for higher levels of thinking, that is, analysis and synthesis (Anderson et al, 2000). Individualized feedback that we realized was missing in our discussion of assessment management systems is the key element of cybercoaching. It is also worth noting that the use of rubrics in this model gives a sense of order to the nature of the qualitative exchange and a system so that results may be used for analysis to certain extent.

Current theory and research in cognitive science have influenced standards of teaching to include the need for students to receive feedback in a timely manner. Providing feedback to students is a process of scaffolding, which reflects the core concept of the 'Zone of Proximal Development', or ZPD (Vygotsky, 1978). From Vygotsky's perspective, learning takes place in a social context. A learner observes and is assisted by a more skilled learning partner. The guide, or helper, detects the need of the learner and renders assistance accordingly. The emphasis on self-regulated learning is a strong positive aspect of this

model as it reinforces the belief that learners are cognitively active participants and key decision makers with regards to their own learning path.

Now to discuss a few concerns related to this model. The first issue that might arise would be how such a method would fit into traditional educational structures where resources are constrained and stereotypes about what a classroom should be exist. It might be difficult to jusify the one to one attention that this model requires and also to be able to prove the benefits of building such an assessment component. Also, matching of cybercoaches to the students would require some logistical work in terms of maintaining databases of updated skill summaries and coordinating availabilities for an additional workload. Even though the general assumption regarding the use of technology is that it indicates a 'mass production' approach to education, the truth is that a single cybercoach through the nature of this model would not be able to work with many students.

Finally it is very significant that both the cybercoach and the learners develop a set of new media literacies (Jenkins, 2009) to be able to function well in this relationship. In fact the cybercoaches need to have pedagogical content knowledge that would help them specifically to share and build knowledge using the cybercoaching model.

Virtual Laboratories

Design is at the heart of engineering, but unfortunately, typically not at the heart of engineering education. Engineering education in the 1990s was oriented towards analysis

rather than design. Unfortunately, there are many reasons why design does not permeate engineering education. Design exercises historically consume tremendous resources (especially time) to create, supervise, and evaluate. In many domains cost and/or safety concerns preclude tackling many design tasks that would be illuminating (e.g., designing power plants or jet engines). The sheer amount of detail work needed to produce a working artifact can be important for integrative exercises, but it detracts from the value of design exercises aimed at teaching just one principle. These factors are leading to an increasing role for software in virtual laboratories that enable students to "build" designs and try them out without expense or danger. Examples of commercial software that can be used this way include MultiSimTM for electronic circuits and Working ModelTM for mechanics (Forbus et al, 1999). When used properly, such programs provide valuable experiences for students.

Virtual Laboratories for the Capstone Engineering Courses

The Virtual Laboratories we are going to discuss here have been developed for the Capstone engineering courses at the Oregon State University in which students have an opportunity to practice engineering. In addition to providing students the opportunity to practice engineering, capstone courses facilitate the development of creative and critical thinking, which are critical in the practice of engineering (Shavelson, 2003). By design, capstone courses are the mechanism by which students apply the core concepts that are critical to their discipline to solve an open-ended problem; this type of activity should enable students to engage in a deeper level of cognition than experienced earlier in their curriculum, which

focuses more on analytical skills.

The two virtual laboratories are the Virtual Chemical Vapor Deposition laboratory and the Virtual BioReactor laboratory (Koretsky, 2008; Koretsky, 2006a; Koretsky, 2006b; Amatore, 2007). These have been built using simulations based on mathematical models implemented on a computer to replicate a physical laboratory.

The virtual laboratories themselves provide useful quantitative data about the student's solution path. A summary of all students' interactions with the equipment and all raw data is available through the instructor interface. A summative evaluation of the performance is also calculated allowing the interactive aspects of their laboratory practice to be analyzed. In the present instructional design, students are required to complete the following: a design memorandum and instructor coaching session that is completed prior to the students performing virtual experiments; a project update memorandum; a project journal; a written report; and an oral presentation. These form records of the student thought processes as they engage in the project.

Discussion

Virtual laboratories can facilitate the creation of microworlds for students to explore in order to discover hidden rules or relationships. They open up a whole new world of assessments and simulations with complex data sets that could never be done through traditional methods. They support situated learning principles (Lave & Wenger 1991) perfectly as learning takes place in the same context in which it is applied and they give us the ability to hone and assess complex cognitive skills.

A fact to consider however, is that typical virtual laboratories do not provide explanations, relying on human instructors and lab assistants to provide the scaffolding necessary for students (Forbus et al, 1999). This means additional efforts at monitoring students who may require support and guidance but are not requesting it. Just like in problems in real life, such programs do not help students understand exactly where and how their assumptions lead to problems. They do not provide coaching or help students tie their results back to the phenomena of the domain, nor do they provide advice on how to improve a student's design. They also do not provide support for assessment administration—instructors still have to grade every aspect of a student's work by hand, a process which is often less convenient for work submitted on-line than on paper.

There has been significant interest in overcoming these limitations by using artificial intelligence techniques to create articulate virtual laboratories, software that has a conceptual understanding of the domain being taught and uses that understanding to scaffold students in design tasks. This is an interesting technology and shows great potential for judging the concrete impact of learning.

Advanced Analytics

Learning analytics is the use of intelligent data, learner-produced data, and analysis models to discover information and social connections for predicting and advising people's learning (Siemens 2010). It is the set of activities a learning organization does that helps it understand how to better train and develop employees.

Learning Analytics relies on some of the concepts employed in web analytics, through tools like Google Analytics, as well as those involved in educational data mining. These analytic approaches try to make sense of learner activity (through clicks, attention/focus heat maps, social network analysis, recommender systems, and so on). Learning analytics is broader, however, in that it is concerned not only with analytics but also with action, curriculum mapping, personalization and adaptation, prediction, intervention, and competency determination.

Learners do generate a significant amount of data that can provide extremely valuable insights for educators. Sources include tweets, facebook updates, logging into a learning management system, or blogposts, learner profiles and other times unintentionally while in the course of daily affairs (or data that is provided by someone else – such as being tagged in Facebook).

With data becoming increasingly intelligent (semantic or linked data), learner data, profile information, and curricular data can be brought together in some form of analysis. A user's data trails and profile, in relation to existing curriculum, can be analyzed and then used as a basis for prediction, intervention, personalization, and adaptation. Just as individuals communicate social intentions through signals well before they actually "think" they make a decision, learners signal success/failure in the learning process through reduced time on task, language of frustration (in LMS forums), long lag periods between logins, and lack of direct engagement with other learners or instructors (Siemens, 2010).

Learning content should be more like computation – a real-time rendering of learning resources and social suggestions based on the profile of a learner, their conceptual understanding of a subject, and their previous experience. Competence (as measured by a degree or certificate) need not be explicitly pursued. Some freely available tools that might be useful for learning analytics applications include Google Analytics, Userfly, Tynt, SNAPP, Piwik, StatzMix.

It has become relatively common within education to use data analysis for tracking and measuring performance at the school, educator, and student levels. Far fewer schools and colleges have taken analytics to the next level. Using advanced analytics for predictive modeling marks a total structural change in education. IBM offers tools, such as SPSS Modeler, that can help schools and colleges improve outcomes—both on the education side and the business side by identifying trends and allowing administrators to make decisions based on patterns and associations found within their data (IBM, 2010).

While many schools have become adept at tracking and measuring student and teacher performance, structured information such as grades, state test scores, and attendance—as well as other information, such as food service data and bus routing schedules—is typically kept in a variety of different repositories. Then there's the high volume of unstructured data that can exist across a campus: information about disciplinary action, parent-teacher reviews, notes from a counselor, surveys of parents or students, or notes from PTA members about different issues affecting a school, to name a few.

Data mining tools such as SPSS Modeler or SAP's BusinessObjects analytics allow institutions to pull together both structured and unstructured data. This is important, because 80 percent of the data produced every day—from email messages to call logs to blogs on the internet—are unstructured (Nastu, 2010).

IBM SPSS Modeler implemented in Tennessee District

The Hamilton County Department of Education (HCDE) oversees nine K-12 school districts in and around Chattanooga, Tenn. Evaluating and improving school performance became a critical task for the districts, owing largely to No Child Left Behind. HCDE officials knew the students in their districts were scoring below state target levels, but it was difficult to understand why—and without that under- standing, it was nearly impossible change the situation. HCDE also had a high dropout rate among high school students, and officials wanted to reduce that number (Nastu, 2010).

Administrators chose IBM's SPSS Modeler and SPSS Statistics software to take a deeper look at student performance by combining data sources and exploring variables beyond what the state reports provided.

Now, HCDE evaluates student performance and keeps students on track earlier in their academic careers by analyzing students' test scores and combining that information with information on student attendance, behavior, parent information, class schedules, and other data.

For example, even before the school year begins, teachers now have a remarkable amount of data on the makeup of their classrooms and on which students might require additional

instruction and focus. A teacher will know in September if the data predicts that a student will not perform well on the early college assessment ACT Explore test, which takes place in November. That student then can be given the extra attention needed to bridge the gap and, ideally, exceed expectations.

This has led to an improvement in test scores, with Hamilton County students performing well above the national average for the ACT Explore test in English, math, reading, science, and overall composite categories for the last three years.

Also using analytics it was identified that the biggest indicator contributing to the high school dropout rate was being overage. Overage students required more attention and suffered from motivational deficit issues and feelings of inferiority. Understanding the high correlation between overage students and dropout rates allows HCDE to be proactive. Officials can identify a student coming into ninth grade who is already 16 or 17 and help the student before he or she gets into trouble, Even earlier in the process, educators can make sure that students—particularly those who have late birthdays—don't get held back more than is absolutely necessary.

Discussion

A key point of discussion as far learner or advanced analytics are concerned is privacy issues and concerns that may come up. There is a need to examine the ethics of using learner generated data in this manner. Particularly because the method by definition pulls information from numerous sources and the learners may not even be aware of the fact that data is being used for 'evaluation' purposes. This is a significant area for smaller

organizations, institutions and schools to explore before they put themselves at risk for expensive litigation.

A significant advantage of advanced analytics software is that perfect data and a fully formed hypothesis are not needed to begin using predictive modeling. We are well aware that education is currently rooted in traditional structures and methods and it is reassuring to learn that these software do not assume a drastic reorganization or change in database practices to be of some use.

Schools that have invested in building out a more robust data offering, and that have begun aggregating student components of data records, test scores, and other information, will benefit from having all this information pulled together. But even schools that haven't had the resources to do that can begin to use predictive analytics.

Another important element in getting started with predictive analysis is a group of people with advanced analytics training. This is a field that requires a pipeline of talented professionals who are essential to the decision making process. While the software itself simplifies the analytics to some extent having a team of experts to implement and maintain this system is essential.

CONCLUSION

This report has looked at some of the emergent practices in evaluating learning using web-based technologies and affordances that are in use today. These tools address a wide range of evaluation needs in innovative ways. While the preceding chapters have introduced you to some of the related literature and the specific technologies that are available, an appropriate way to conclude might be to link this knowledge to a practice context through scenarios created for this purpose. These scenarios show a few typical challenges and problem situations that the tools and working models that we have discussed at length seek to address. I intend for the readers to use these scenarios as intellectual puzzles and the analysis only as a starting point for discussion and thought leading to even better solutions.

Scenario 1

A certain university has a large continuing education center which is designed to provide a wide range of training and educational opportunities for individuals getting started in the workplace, furthering their current careers, or working to meet licensure/certification requirements. The majority of their training has been in the instructor-led category but they have recently ventured into the on-demand web based training space as well. An analysis of their current processes shows that they have been duplicating a lot of work in terms of assessment items because they do not have an organized system for maintaining and reusing questions created. Answer sheets for tests are scored by hand and given the large number of enrollments this is an administrative nightmare. Since they are moving

into a blended format for their offerings they need to find a solution that would help them deliver assessments both in a web-based as well as an instructor-led format. Another concern for them is that some of the testing they perform is high-stakes and they would need measures to counter any breaches in this regard. In terms of their web-based offerings, they require the ability to use multiple question types and interoperability with other systems.

Analysis

A good option might be for the department to look at assessment management systems (Respondus, QuestionMark Perception) available in the market that are a good fit for their requirements. These solutions are usually available in versions with differing levels of features enabled and appropriately priced. Once they have created a detailed document with an outline of their specific needs and requirements they can work with the vendors to work a solution that suits both their needs and budget requirements. The two products mentioned here have features like multiple question types, randomization of questions, easy print templates and options for the instructor led components which seem like a good fit with this organization's requirements. A domain lock is also available for an extra fee to secure their high stakes assessments.

An interesting concern here might be the existence of legacy data (older assessment items) that needs to be incorporated into the system. A helpful feature most such tools offer is the ability to load assessment items into the system in word document formats so

that users who are not well versed with the tools would not have problems. Older items that are in word document formats might be entered into the database using such templates.

Scenario 2

The content development team of an E-learning firm has been working on a range of e-learning products for the petrochemicals industry. They have proceeded through all the recommended phases of the development process (market survey, needs analysis, creation of design documents, development) and are very close to launching their products. As a part of preparing for the launch, the company sent out a team to market their products to groups of potential clients through presentations demonstrating and introducing this new range to them. One of the questions raised was about what measures they had taken to ensure that the modules was appropriate for the learner audiences they were for. The customers were also concerned about whether usability considerations had been incorporated. It was all well to start out with a detailed learner analysis but did the company have any inputs on whether the learners would see value in the finished products. The challenge now lies in collecting these inputs in an efficient and cost effective manner.

Analysis

A possible course of action might be for the company to identify and obtain contact information of a list of people who might be willing to participate in a formative

evaluation exercise and who are representative of the target learner group. Once that is done and they could choose a method to collect the data based on their time constraints, requirements and technical capabilities. One option would be to use one of several online survey tools available (free versions as well as premium versions are available for a fee) to help them perform the formative evaluation that they need to improve or validate the quality of their products. This would make it easy to administer and track responses and reach out to large numbers of people. Most survey tools also provide significant data analysis features that would make it easier to draw conclusions from the responses received.

Another option, since this is a formative evaluation exercise for a web based product might be to use a freely available user analytics tool like userfly which records the user's actions as a video for later analysis about usability concerns and whether the course is being received the way that it was intended. This has the advantage of being able to observe the viewers while they are performing the tasks as opposed receiving survey responses that might be tempered with the respondent's personal biases. It must be noted however, that if there are a very large number of respondents this could be very tedious and time consuming in terms of analysis.

Scenario 3

The global training team heads of a rapidly growing Insurance company recently met to discuss their strategic vision for employee development initiatives and one of the areas of concern that emerged from those discussions was the need for leadership development

among new team leaders across all functions. The first thing they needed to do was to evaluate where their fairly large learner audience of team leaders stood in terms of leadership competencies. They discussed the huge expenditures in terms of time and money that this might involve if they had to have development centers organized for them at the numerous locations that they were spread across. They are researching easier and more efficient ways of performing this exercise without disrupting work (for the participant employees) and making a huge investment. The quality of input received and the psychometric soundness of any methods they use are also of utmost importance to them.

Analysis

One of the solutions this company might possibly use could be an online 360 degree feedback tool to help them with the administrative hassles of testing large learner audiences on leadership competencies. It would be important of course to do some preliminary research about the service providers and their credentials. The primary reason being that it would be unfair to the employees being assessed if these measures were not taken and learning initiatives would not have their desired impact. Another concern would be that assessing leadership skills is an inexact science and if there is room for error in a system affecting an employee's careers significantly the organization might have to face unpleasantness and legal issues in the future.

It might also be useful for the company to incorporate aspects of the cybercoaching model in their training approach for leadership skills in the longer term. The principles of

immediacy of feedback from the coach, self regulation and participation in the design of rubrics seem well suited to a skill set of this nature.

REFERENCES

ADL.(2010). *SCORM*. Retrieved from http://www.adlnet.gov/Technologies/scorm/default.aspx

AICC.(2010). *About the AICC.* Retrieved from http://www.aicc.org

Alkin, M. (Ed.). (2004). *Evaluation roots: Tracing theorists' views and influences.* Thousand Oaks, CA: Sage.

Anderson, L., Krathwohl, D., Airaisian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J. & Wittrock, M. (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Boston: Allyn & Bacon.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W.H. Freeman.

Bartlett, T.(2004). Taking Control of the Classroom. *Chronicle of Higher Education, 51*(4), A8-A9

Bergstrom, B., Fryer, J., & Norris, J. (2005). Defining online assessment for the adult learning market. In Hricko, M., & Howell, S.L. (2005). Online Assessment and Measurement: Foundations And Challenges (pp. 46-66) Information Science Publishing: Hershey.

Bower, M. (2005). Online assessment feedback: Competitive, individualistic, or preferred form *Journal of Computers in Mathematics and Science Teaching, 24*(2), 121-147.

Brindley, C., & Scoffield, S. (1998). Peer assessment in undergraduate programs, *Teaching in Higher Education, 3*(1), 79-89.

Carey, L.M. (2001). *Measuring and evaluating school learning* (3rd ed.). Boston: Allyn and Bacon

Cassady, J., Budenz-Anders, J., Pavlechko, G., & Mock, W (2001, April). The effects of internet-based formative and summative assessment on test anxiety, perceptions of threat, and achievement. Paper presented at the annual meeting of *the American Educational Research Association*, Seattle, WA.

Chang, W., Hsu, H., Smith, T.K.& Wang, C.(2004). Enhancing SCORM metadata for assessment authoring in e-Learning. *Journal of Computer Assisted Learning, 20*(4), 305-316

Chatzopoulou, D. & Economides, A.(2010). Adaptive assessment of student's knowledge in programming courses. *Journal of Computer Assisted Learning, 26*(4), 258-269

Clark, K. (2004). What can I say besides sound it out? *Reading Teacher, 57*, 440-450

Dede et al. (2009). A Research Agenda for Online Teacher Professional Development. *Journal of Teacher Education*, 60(1), 8-19

Dermo, J. (2009). E-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology, 40*(2), 203-214

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review, *Studies in Higher Education, 24*(3), 331-350.

Duda, M. D. & Nobile, J.L.(2010). The Fallacy of Online Surveys: No Data Are Better Than Bad Data. *Human Dimensions of Wildlife*, *15*(1), 55-64

ETS. (2010). *The GRE® revised General Test is Coming in August 2011.* Retrieved from http://www.ets.org/gre/revised_general/know

Ferrao, M. (2010). E-assessment within the Bologna paradigm: evidence from Portugal. *Assessment & Evaluation in Higher Education*, *35*(7), 819-830.

Fetterman, D.M. (2001). *Foundations of empowerment evaluation.* Thousand Oaks, CA: Sage.

Fluckiger, J., Vigil, Y. T., Pasco, R. & Danielson, K. (2010). Formative Feedback: Involving Students as Partners in Assessment to Enhance Learning. *College Teaching, 58*(4), 136-140

Forbus, K.,Whalley, P.B., Everett, J.O., Ureel, L., Brokowski, M., Baher J., Kuehne, S.E.,(1999). CyclePad: An articulate virtual laboratory for engineering thermodynamics, *Artificial Intelligence 114 (1999,)* 297–347

Friesen, N. & Koohang, A (2005). Interoperability and Learning Objects: *An Overview of E-Learning Standardization* 1, 23-31

Heron, J. (1981). Assessment revisited. In D. J. Boud (Ed.), *Developing Student Autonomy in Learning*. London: Kogan Page.

Harvey, J., & Mogey, N. (1999). Pragmatic issues when integrating technology into the assessment of students. *Computer-assisted assessment in higher education* (pp. 7-19). London: Kogan Page Limited.

IBM. (2010). *IBM SPSS Modeler Professional.* Retrieved from http://www.spss.com/software/modeling/modeler-pro/

Jenkins,Henry. (2006) Confronting the Challenges of Participatory Culture: Media Education for the 21st Century.

Kirkpatrick, D.L. & Kirkpatrick, J.D.(2006). *Evaluating Training Programs*, Berret-Koehler Publishers, Inc. CA: San Francisco

Koretsky, M.D., Amatore, D., Barnes, C., & Kimura, S. (2008). Enhancement of Student Learning in Experimental Design using a Virtual Laboratory. *IEEE Trans. Ed.*, 51, 76-85.

Koretsky, M.D., Barnes, C., Amatore, D., & Kimura, S. (2006a). Experiential Learning of Design of Experiments Using a Virtual CVD Reactor. *Proc. Ann. ASEE Conf. & Expo.*, Chicago. Koretsky, M.D., Amatore, D., Barnes, C., & Kimura, S. (2006b). The Virtual CVD Learning Platform. *36th ASEE/IEEE*

Lifelong Learning Market Report. (2004). *E-Learning Standards Groups ADL And AICC Work To Address Content Integration Issues. 9*(17),1-5

Lin, C.-J. & Spray, J.A. (2000). *Effects of item-selection criteria on classification testing with the sequential probability ratio test.* (Research Report 2000-8). Iowa City, IA: ACT, Inc.

Looney, Marilyin (2008). Using an Online Survey Tool to Enhance Teaching. *Measurement in Physical Education and Exercise Science,* 12: 113–121

Mazzolini, M. & Maddison, S. (2007). When to jump in: The role of the instructor in online discussion forums. *Computers & Education, 49*(2), 193-213

McCann, A.L. (2010). Factors affecting the adoption of an e-assessment system. *Assessment & Evaluation in Higher Education, 35*(7), 799-818

McFarlane, A. (2003). Assessment for the digital age. *Assessment in Education: Principles, Policy & Practice*, *10,* 261-266

Mcleod, Scott. (2009). *Teaching and Learning in an era of Disruptive innovation.[PowerPoint slides]*. Retrieved from http://www.3dwriting.com/mcleod/

Morisano, D. & Shore, B. (2010). Can Personal Goal Setting Tap the Potential of the Gifted Underachiever? *Roeper Review, 32*(4), 249-258

Mowl, G., & Pain, R. (1995). Using self and peer assessment to improve students' essay writing: A case study from geography, *Innovation in Education and Training International*, *32*(4), 324-335.

Nastu, J. (2010). *Advanced Analytics: Helping educators approach the ideal.* Retrieved from http://www.eschoolnews.com/wp-content/uploads/2010/09/eSMSep2010_ADVANCED-ANALYTICS.pdf

Northcote, M. (2002). Online assessment: Friend, foe or fix? *British Journal of Educational Technology, 33*(5), 623-625.

Questionmark Corporation(2010) *Questionmark Case Study: Penn State University.* Retrieved from http:// www.questionmark.com/us/casestudies/penn_state.aspx

Robles, M., & Brathen, S. (2002). Online assessment techniques. *Delta Pi Epsilon Journal, 44*(1), 39-49.

Rovai, A. P. (2000). Online and traditional assessment: What is the difference? International and Higher Education, *3*(3),141-151.

Searby, M., & Ewers, T. (1997). An evaluation of the use of peer assessment in higher education: A case study in the School of Music, Kingston University, *Assessment and Evaluation in Higher Education, 22*(4), 371-383.

Shavelson, R., Ruiz-Primo, M.A., Li, M. & Ayala, C.C. (2003). Evaluating New Approaches to Assessing Learning. *Center for the Study of Evaluation Report*, *UCLA*.

Siemens, G. (2010). *What are Learning Analytics?.* Retrieved from http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

Singh, A., Taneja, A. & Mangalaraj, G. (2009). Creating Online Surveys: Some Wisdom from the Trenches Tutorial. *IEEE Transactions on Professional Communication, 52*(2), 197-212

Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a Learning Environment by Using Self-Peer- and Co- Assessment. *Learning Environments Research, 1*, 293-319.

Thissen, D., & Mislevy, R.J. (2000). Testing Algorithms. In Wainer, H. (Ed.) Computerized Adaptive Testing: A Primer. Mahwah, NJ: Lawrence Erlbaum Associates

Tierney, R.J., Carter, M.A., & Desai, L.E. (1991). "Portfolio assessment in the reading-writing classroom." Norwood, MA: Christopher Gordon

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students, *Assessment and Evaluation in Higher Education, 25*(2), 146-169.

Torrance, H. (1997). Assessment, accountability and standards, using assessment to control the reform of schooling. In: Halsey, A.H., Lauder, H., Brown, P. & Wells, A.S., (Eds) *Education, Culture, Economy and Society* (Oxford, Oxford University Press).

Velan, G., Kumar, R., Dziegielewski, M & Wakefield, D.(2002). Web-based self-assessments in pathology with Questionmark Perception. *Pathology, 34(*3), 282-284

Voogt, J & Knezek, G. (2008) *International Handbook of Information Technology in Primary and Secondary Education ,* Springer. NY: New York

Vygotsky, L.S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wainer, H., & Mislevy, R.J. (2000). *Item response theory, calibration, and estimation.* In Wainer, H. (Ed.) Computerized Adaptive Testing: A Primer. Mahwah, NJ: Lawrence Erlbaum Associates.

Walker, C. (1998, January 19). Pair takes training online. *PC Week, 15*(3), 65

Walton, S. (2005). *The eVIVA project: Using e-portfolio in the classroom*. BETT, January 2005. Retrieved from www.qca.org.uk/downloads/10359_eviva_bett_2005.pdf

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. Journal of Educational Measurement, 21, 361-375.

Wills, G.B., Davis, H.C., Gilbert, L., Hare, J., Howard, Y., Jeyes, S., Millard, D & Sherratt, R. (2009). Delivery of QTIv2 question types. *Assessment & Evaluation in Higher Education, 34*(3), 353-366

Williams, D.D., Hricko, M., & Howell, S. L. (2006). *Online Assessment and Measurement and Evaluation: Emerging Practices*, Information Science Publishing. PA: Hershey.

VITA

Hemangini Dutt Majumder has three years of work experience (including project management experience) in the field of corporate learning and development in the Oil and Gas, Insurance, IT consulting and Healthcare sectors. She has worked on projects presenting great variety in terms of content (soft skills as well as more technical knowledge domains) and learners (ranging from entry-level sales force to senior management). Her educational qualifications include an MBA with a specialization in Human Resources and a Master's degree in Instructional Technology (graduating in Dec 2010). She enjoys her field of work, finding it both creatively fulfilling and challenging.

Permanent address: J-1/2 Himali Society Erandwane Pune 411004 India.

This report was typed by the author.