

Andrews University

School of Education

EVALUATION AND ASSESSMENT:
FINDING QUALITY IN DIVERSITY

In Partial Fulfillment

Of the Requirements for the Leadership Program

LEAD 756

by

Janine Lim

July 2009

This file is copyrighted (2010) and a part of my Leadership Portfolio found at <http://www.janinelim.com/leadportfolio>. It is shared for your learning use only.

Introduction

In this reflection paper, I consider my current practices of assessment and evaluation in light of the knowledge base I have reviewed. In my exploration, I found that my constructivist beliefs about knowledge and learning have profoundly affected my views on this topic. So this reflection will begin with a comparison of assessment and evaluation, then move into my practices of assessment and evaluation, and finish with

I chose the tagline of “finding quality in diversity” for this competency. This picture is a selected close up from my “out on a limb” theme picture. I chose this picture because each leaf is different in size, shape, and lines, yet they are clearly still from the same tree. In my constructivist view of



assessment, participants in my classes and workshops may create varying products through varying processes, yet each of them are quality learning experiences. In addition, my work in assessment goes out on a limb because of its non-traditional view. My evaluation work is just beginning, and the wide field of evaluation is opening up to me. In this arena as well, I look for quality among diverse experiences in the programs I am involved with. While my evaluation and assessment may be “out on a limb”, it still informs my programs and plans for future work.

Comparing Assessment and Evaluation

When I began this study, I found the difference between assessment and evaluation. People in evaluation use the word assessment (i.e. needs assessment), and

people working in assessment use the word evaluate (i.e. use assessment to evaluate student learning). Walvoord (2004) suggests that direct assessment measures student work and indirect assessment measures overall program goals and values. However, Baehr (2009) clarifies the purpose of the two. The purpose of assessment is to improve the quality of performances in the future; whereas the purpose of evaluation is to determine the quality of the present performance. Assessment is used to improve performance, whereas evaluation is used to make judgments. Another view is that “assessment relates to the performance of individuals, while evaluation relates to systems or processes” (Shephard, 2009, p. 395).

McNeil, Newman, and Steinhauser (2005) suggest that evaluation is a systematic process that includes determining standards for making the judgment, collecting information, and then applying the standards to determine quality. Evaluation is much closer to research, as evaluators discuss and debate various research designs for effective evaluation (Cook, 2003; Fitzpatrick et al., 2004; Haertel & Means, 2003; Heineke & Blasi, 2001). Evaluation could be considered applied research, with the goal of finding results that can be easily and effectively implemented (McNeil et al., 2005).

I use assessment in my teaching: for online classes, and for face to face workshops. I use evaluation to make decisions and plans about my videoconference program, smaller videoconference projects, and grant implementations. Since I am not a professional evaluator, my evaluation work does not rise to the complexity that I have learned about in my reading (Fitzpatrick et al., 2004; Love, 1991; Patton, 1997, 2002; Worthen et al., 1997). Yet, as a program director, understanding the process of evaluation will improve my practice and willingness to gain and receive feedback on my programs (McNeil et al., 2005).

Assessment and Constructivism

As I began my study of assessment, I found a discomfort with assessment tools such as multiple choice questions, quizzes, tests, exams, and so forth. I do not use these assessment tools in my courses, and I share educators' frustrations with high stakes testing (Janesick, 2006). This study has forced me to examine why.

My beliefs about learning and knowledge are strongly constructivist. I believe learning is active; the learner has prior knowledge, and that the learner should be taking responsibility of their learning (Freedman, 1998). In this learning context, students apply their knowledge in real-world scenarios and exhibit their knowledge by applying the knowledge. Some constructivists disagree on how assessment should occur. Should educators assess the process (Jonassen, 1991) or the product (Cole, 1992) or some combination of the two? Some suggest that assessment occurs when the teacher elicits prior knowledge, during dialogue and inquiry, and during the process of the learning activity (NCREL, 2004). In this scenario, assessment is student centered and outcome based and includes self and peer evaluation. The assessment is focused on the process of constructing knowledge and is context dependent (Matusevich, 1995). This view of learning definitely affects my preference for authentic assessments such as demonstrations, observations, journal writing, and portfolios (Janesick, 2006). Janesick suggests that authentic assessment requires a quality performance or product, is connected to the student's world, is complex and continues with multiple tasks, and provides continually recurring feedback. A recent study confirmed that students preferred an open book open web exam, designed to require solving a real-world problem, and that this type of final exam is more compatible with constructivist learning theory (Williams

& Wong, 2009). These principles provide the foundation for assessment in my workshops and online classes.

Thinking About Assessment in Online Classes

Over the last 10 years, I have written and taught nine online courses in instructional technology. These courses are offered for graduate credit through local universities. In addition, I teach many short face to face workshops and an occasional week long graduate credit face to face workshop. Each of the graduate credit courses are set up with assignments for readings and application of the work. Participants read articles and discuss them in the online discussion; and then do assignments that require application of their learning. The assessment in these assignments is authentic and based on their writing, their interaction with others, and their performance on a product applying their learning (Janesick, 2006). Teachers attending a technology class tend to have a wide range of skill level. Some participants receive awareness knowledge from the course; others are ready to implement when they are finished. My goal is that each participant moves forward along a continuum of learning; and because I provide a somewhat overwhelming amount of options and choices, I know they always make progress in their learning. This is my current practice.

As part of my work in this competency, I read several articles about assessment in the online environment, or “e-assessment” (Whitelock, 2009) while thinking about my own online courses.

Some of the assessment strategies were geared towards knowledge acquisition content in the sciences and math. These types of assessment strategies would not work in my courses where participants are applying their technology knowledge to their curriculum and classroom practice. Still, as I have taught courses on how to teach online,

these articles contributed to my knowledge of the assessment tools available which will help me more appropriately advise others in teaching online. This includes strategies such as regular online low-mark quizzes for content such as mathematics (Angus & Watson, 2009; Marriott, 2009), an Open Mark open source tool for grading and giving students feedback on short answer questions using natural language (Jordan & Mitchell, 2009). In addition, students have generally positive attitudes towards e-assessments that supplement face to face classes (Dermo, 2009).

Other assessment articles were intriguing in light of the AU Leadership Program. The netfolio concept (Barbera, 2009) provides for students to see and comment on each other's artifacts as they work on a particular standard or section of their portfolio. This reminded me of the Leadership and Learning Group process and the ways that we learn from each other. The portfolios in the study were created in a research class, and the principles could apply to the dissertation proposal class as well. Peer assessment and feedback allows for more reflection, self-evaluation, and a higher cognitive engagement on the topic. Another portfolio study suggested that students need to learn how to peer-assess and give appropriate feedback (Chang & Tseng, 2009), which may suggest future webinars for the Leadership faculty. A study of undergraduate dissertations suggested a conversation theory framework for feedback between advisors and their students (Heinze & Heinze, 2009). An online peer-moderating marking system could be used to review the group process and functioning in Leadership and Learning Groups (Loddington et al., 2009).

Still other articles confirmed or challenged the processes I use in my online courses. Peer feedback in completing a task provides greater engagement and learning for students (Barbera, 2009), which is encouraging because participants in my classes give

each other weekly feedback. In another article, Draper (2009b) described the six ways that students interpret feedback. This knowledge can inform my interaction with students in my classes, to ensure a better match between my intent and the students' interpretation. Another study focused on the use of a dynamic frequently asked questions tool, which allow students to ask questions anonymously on the content or process of the course (Ng'ambi & Brown, 2009). The feedback was very useful to the instructor, who could then make modifications to the instruction based on the questions of the students. I am thinking of trying an anonymous question and answer section for my online classes, although it may not be as necessary with adult learners with small class sizes than for undergraduate classes with hundreds of students.

Finally, some articles sparked my imagination for potential assessment for learning activities. Draper (2009a) suggested six ways to use electronic voting systems or classroom response systems to engage students in learning, all of which require peer interaction.

1. Assertion–reason questions, which can be and have been used with EVS.
2. Taking an MCQ and having the learner generate reasons, for and against each response option, rather than simply ticking one. (This is usually done on paper as a private revision technique.)
3. Confidence-based marking, which is normally delivered by ICT and could be done with EVS with some (but not all) software.
4. Mazur's method of using brain-teasers to prompt peer discussion, which is routinely done using EVS.
5. Having students create MCQs as part of presentations using EVS.
6. Having students create MCQs for use in tests that may be administered either by using EVS or on paper (Draper, 2009a).

These suggestions fit well with my beliefs on the social construction of knowledge. Some of our local schools use these classroom response systems, and these ideas will assist me in providing resources and ideas to our local schools.

Thus, a review of e-assessment articles confirms my current practices and provides ideas to share with others.

Thinking About Evaluation and My Practice

"Learning is experience. Everything else is just information." - Albert Einstein

My study of evaluation was very intimidating to me. My simple little evaluation forms for my programs and projects are so small compared to the detailed evaluation and research designs I reviewed (Haertel & Means, 2003; Heineke & Blasi, 2001). While I have experienced, once, an external professional evaluator evaluating a grant program (See artifact B.ii), I have never even considered completing an evaluation as complicated as I have been reading about (Fitzpatrick et al., 2004; Patton, 2002). Certainly my awareness of the possibilities have grown, and I will be more educated as I work with external evaluators in the future.

McNeil et. al (2005) provide an interesting discussion of how the program director should be involved in evaluation. While my title does not include "director", for most of my programs and projects I am essentially the director. So the suggestions are particularly useful. The program director should see the evaluation not as a negative judgment, but as a valuable tool to increase the likelihood of success of the program. The program director should realize the others will be involved in the evaluation process and should take an active role, assuming some responsibility for many of the evaluation tasks. If the program director is committed to the evaluation, she will be involved with the evaluation before, during, and after the program is conducted and the evaluation will then be used often for improvement. "When the program director takes an active role in the evaluation, the program director focuses more on evaluation information than on anecdotal information" (McNeil et al., 2005, p. 16).

In addition, the evaluator's focus is different than the program director (McNeil et al., 2005) and this may provide a place for me to try to think more like an evaluator and less like a program director when doing my own program evaluations. Program directors are usually more focused on specific elements; whereas evaluators focus on the entire program. Program directors are interested in evaluating the components they can control; evaluators include additional components not under the program director's control. Program directors are interested in evaluating just the program and those outcomes that can be attributed to it; evaluators want to see how the program fits with the regular program. Finally program directors are more often interested in non-quantitative procedures, while evaluators are more likely to use qualitative and quantitative procedures. To improve my practice, I can think more about the program as a whole, all components of the program, and how it fits within the schools. I can and should also broaden the type of data that I collect on my programs.

General Evaluation Model and Improving Practice

To consider my current evaluation practices, I have chosen the General Evaluation Model (McNeil et al., 2005) as it was the clearest method and made the most sense to me. In addition, the authors suggest that this model can be used by an individual instructor in a course, up to evaluating departments and organizations. The model has five stages: (1) needs assessment, (2) baseline, (3) procedures to achieve objectives, (4) program implementation assessment, and (5) post assessment.

In considering the General Evaluation Model, I will think about three main areas: my full videoconferencing program, the Read Around the Planet project, and my smaller videoconference projects. At Berrien Regional Education Service Agency, I coordinate and manage the videoconferencing program for the schools. The Read Around the Planet

project is organized by TWICE, Michigan's K12 videoconference organization, and matches over 1500 classrooms annually to celebrate reading via videoconference. Finally, as part of my work for Berrien RESA, I run several projects for the schools. One of them is called MysteryQuest, where six classes around the country present geography clues and the other classes research to determine what the other class presented. This project and the others like it have an evaluation form that is in the online survey tool, Zoomerang. However some of my other projects do not have an evaluation and probably should. In addition, we offer quite a range of programs interviewing authors and specialists that also do not currently include any evaluation. With these projects in mind, I will now review the five stages to the General Evaluation Model in light of these programs.

Needs Assessment

The needs assessment discovers why the program is needed. The first step is to identify the stakeholders in the evaluation, such as the program director, the program staff, the parents, participants, community members, and funding agency. The next step is to identify the program areas to evaluate. This may be areas that are required for the program, new or questionable components, and certain areas of focus. Next the sources of information should be determined, possibly including testing results, attitudes, observations, existing records. The needs assessment instrument is developed next, and the easiest way to do this is to modify an existing validated needs assessment. Finally the needs assessment is conducted and a report written (McNeil et al., 2005).

When the videoconference program began in Berrien and Cass counties, my supervisor wrote a grant to get videoconferencing in the high schools. The program started with a vision, but did not include a needs assessment. In 2003, I noticed that the

elementary schools appreciated the possibilities of videoconferencing, but found it difficult to travel to the high school to access videoconferencing. I sent a small survey to teachers, and found that 79% of the teachers who responded said they would participate in videoconferencing if they had access in their building. This needs assessment data was used to write the grant, which we received and implemented 2006-2009. Now that the grant is over, it may be time to conduct another needs assessment to determine other barriers to using videoconferencing or what other needs the school districts have. I currently listen to my coordinators to see what they need, but this is informal and not all of them regularly talk to me. A new assessment could discover the needs of teachers, of principals, and of the videoconference coordinators who promote and support videoconferencing.

The other projects, Read Around the Planet and MysteryQuest, have never had needs assessments. A needs assessment for Read Around the Planet may not be desired, as the program has exploded to be almost unmanageable. I am not sure what we would want to find out and if we would want to know the answer. Further discussion with someone with more evaluation experience might provide some answers. A needs assessment for MysteryQuest might be appropriate for my middle school geography teachers. In the first years of the program, many of my local teachers participated. Now fewer of them participate, and a needs assessment may discover the reason why and suggest potential solutions. In addition, a general needs assessment for teachers may suggest future projects to create that match their curriculum needs.

Baseline

The next step in the General Evaluation Model is to collect baseline data. The information to measure in the baseline stage is determined by the needs assessment. This

data may include information about individual characteristics, such as student motivation, achievement or staff commitment to the program. Another area are the context information such as socioeconomic status. A third area is relevant to the task or program and the objectives related to it. Information can be collected from the recipients of the program, the staff, community members, and existing records. This data should be collected as close to the beginning of implementation as possible. In the baseline stage, the expected value of the program should be determined as well. What is the expected gain in motivation or student scores or another measure? In addition, a comparison group should be found if possible. This group would receive “standard treatment” and in theory your program should be better than the current standard treatment. The authors suggest the term “comparison” instead of “control” (Cook, 2003) because it is very difficult to find a group that mirrors the group receiving the new program. Other comparison methods could include pre and post assessment, and using the previous year’s cohort, or referencing national or state norms. A final method suggested by the authors is to test all students at baseline. Those at or above the cutoff on the test receive the regular program, those scoring below receive the regular program and the compensatory program. A prediction equation can be created for the compensatory students using the regular student scores, and then this is compared to their actual scores to yield a measure of the project impact (McNeil et al., 1996; McNeil et al., 2005). Finally, the baseline should be administered, data collected, cleaned, and analyzed, the report written. If necessary, the program objectives should be adjusted based on the data collected, and then the program objectives should be shared with the stakeholders.

None of the projects or grant implementations I have been involved with have ever collected baseline data. Unfortunately, while our constituency wants evidence

(Creighton, 2007), there is a belief in some educational technology circles that it is impossible to effectively evaluate the implementation of technology due to the dizzying array of new technologies, the rapid expansion of functions in software, and the quick obsolescence of technology tools (Baker & Herman, 2003). When I consider my own work, I wonder what baseline data would be collected. My schools range in using videoconference 5 – 70 times a year, with most teachers doing one videoconference and some doing up to 12 in a year. Teachers participating in Read Around the Planet have their students interact with another classroom for about an hour. The MysteryQuest projects are a unit of instruction about two weeks in preparation and 1 to 2 hours of videoconference. Do any of these make enough of an impact to be measured? One could ask, if that is so, why do videoconferencing? Teachers comment above the value that students find in using videoconferencing. Is that measurable? These questions are troubling and frustrating, and further research and reading will hopefully provide some answers. It may be that I need to pick a program and attempt to follow this process to see what results occur.

Procedures to Achieve Objectives

The third stage in the General Evaluation Model is to develop procedures to achieve objectives. This begins by developing a program planning team, having them review and commit to the objectives from stage two, and deciding on the scope and plans of the new program. The planning team must focus on the objects and determine the new program based on those objectives. The next step is to train the staff. The delivery of training includes identification of needed training, systematic devising of the training, and appropriate delivery of training. The program director should ensure that the staff are actually attending the training and that they understand what was provided. Most staff

evaluations are affective, but it is important to check for actual understanding. The program director should ensure that the staff understand what parts of the old program will be discarded and how the new program fits into the old program. The program director should make sure the trainer knows how to train adults and that the training is integrated with previous training. Evaluation tools should be identified or developed to assess if the staff understand the training, preferably given later to determine if the staff used what they learned. This evaluation should then be shared with the participants, the trainer, and the program director. Sharing with the participants is uncommon, but useful to help them see the value of completing the evaluation and to see overall how others viewed the training (McNeil et al., 2005).

There is no training for Read Around the Planet or MysteryQuest and similar projects, so this stage does not apply to those projects. However, my overall videoconferencing program has many different training opportunities. Some workshops have evaluation, some that are only one hour do not. My online courses, two of which address using videoconferencing in the classroom, all have evaluations. The closest that I have come to meeting the requirements described are in the professional development and end of year evaluations for the recent grant implementation. This data asked the participants before the workshop, after the workshop, and at the end of the year their comfort level with using videoconferencing on a Concerns Based Adoption Model 6 level scale. The participants moved one full point on the scale from before and after the training and another full score after the end of the year. However, collecting and analyzing this type of data is not my common practice. In fact, most of my workshops either do not have an evaluation or the data is qualitative and not comparable from workshop to workshop. Clearly it would be very beneficial to design a consistent

workshop and course evaluation form that measures whether the participants understand what they learned, plan to implement it, and also determines if they actually did implement it. This type of evaluation should probably be done at the organizational level for all of our workshops and courses.

Implementation Assessment

The fourth stage of the General Evaluation Model is program implementation assessment. The question in this assessment is whether the program is being implemented as planned. The sources of information for this assessment are the program director, the staff development trainer, and the staff implementing the program. The program implementation plan contains the dates the events will occur, the events that will be evaluated, and how to do the evaluation of the events. Instruments for evaluation should be identified and developed if necessary. Stakeholders should be informed periodically. Context information should be collected that describes the setting, staff and participants where the program is delivered. Input information, the resources that were included in the program and when they were provided should also be collected. Process information should also be collected, and this includes the training of staff and how the staff provides the services to the services. A staff development workshop should include a list of the staff development topics, some anonymous demographic information, a Likert scale on how well the topic was covered, an assessment of the likelihood of implementing techniques, and an invitation to identify what will be implemented in the next two weeks. Delivery of services should also be assessed, most often by the evaluator observing classroom teachers but also could be via interviews with principals and possibly parents. Information from this evaluation should be used to adjust the program as soon as

possible. During this process, it is helpful to check that the program is still connected to the objectives and communicate this to those involved (McNeil et al., 2005).

Again, because participants generally sign up for Read Around the Planet and MysteryQuest, participate and leave, there isn't really training or opportunity to do this type of implementation evaluation. Whether or not the teachers implemented the program usually is shown when they show up (or not) to the videoconference. It may be helpful to collect data from those who do not show up; however usually they are busy or some family emergency has arisen and a form to complete would be unwelcome. On the other hand, this step is very appropriate for my videoconference program. An area that I don't usually collect information is the context information. It may be helpful to understand more about the context in order to understand why some schools are using their videoconference system more than others. In addition, this stage provides additional suggestions to improve the evaluation of professional development. Our videoconference calendars show how many events are scheduled, so schools and district level personnel can easily see how well teachers are using videoconferencing throughout the year. In addition, I use this data to periodically call or email program offers to schools that are under-utilizing videoconferencing. We have had evaluation of videoconferences with content providers in the past, but due to some database problems those evaluations are currently not occurring. However, it is on my list to bring this evaluation source back up in the next school year.

Post Assessment

Most programs are funded or implemented annually, so this cycle of stages should occur annually. The post assessment occurs at the end of the school year. The question to answer is how much effect has the program had? First, the evaluation team should

develop a post assessment plan focusing on the effectiveness of the objectives as compared to the baseline data. Next instruments should be identified or developed. The information should then be collected ideally under the same conditions as the baseline and with the same sample participants that were included in the baseline. The same analysis procedures used with the baseline data should be used at the post assessment as well. In addition, ideally the post assessment will be used as the baseline for the next cycle of evaluation. Finally, the end-of-cycle report should include an executive summary, an overview of the program, a listing of the major activities and components of the program, the results of each objective, and a summary of the results. The reports should be disseminated and reflected on for further improvements in the program.

The first thing that strikes me with this level is the question to be answered. What is the effect of the program? My current practice of counting events is not sufficient. Is it possible to determine the effect of Read Around the Planet, MysteryQuest, and my videoconference program? This comes back to my earlier conundrum of what is measurable at the baseline stage. In addition, the format of the end-of-cycle report suggests improvement and addition to my current videoconference annual reports.

Reflection

It is clear that my current practice appropriately includes some aspects of evaluation. The General Evaluation Model could be applied in a broader manner on my overall videoconferencing program and future grant implementations. To do so, though, I must resolve the question of what could determine effectiveness. Rumberger (Rumberger, 2003) acknowledges the challenges inherent in determining student impact with performance-based assessments on constructivist learning applications of educational technology. I am still not clear on this, but hopefully in further discussions

with my videoconferencing colleagues, and further reading on evaluation of educational technology an answer can be found.

Future Learning

As seems to be common in working on a competency in the AU Leadership program, I have again just scratched the surface of the knowledge base for assessment and evaluation. I realize the depth of both topics, and the breadth of support and resources available for each (Stufflebeam & Wingate, 2005). My awareness has been raised to a new level, and I have begun to consider the ways to improve my practice based on what I have learned.

I only reviewed in depth one evaluation model, but there are many more. A few of them are now briefly described. Hense (2009) suggests the use of a logic model to consider input variables, process variables and program effects or outcome. Love's (1991) internal evaluation model includes the initial contact and request for evaluation, scouting the problem, diagnosing the program, contracting duties, managing the evaluation, reporting, and monitoring for future progress. Another type is utilization-focused evaluation, designed with the goal that the evaluation will provide practical useful information and actually be used (Patton, 1997). Evaluations may be formative or summative, internal or external, focused on process or outcome, objectives focused or goal-free, management or consumer oriented, expertise or participant oriented (Fitzpatrick et al., 2004). Evaluations of educational technology could include longitudinal designs or studies to determine cumulative impacts (Haertel & Means, 2003). While I have yet to fully implement one of these evaluation models and so feel that I do not "know" them by experience, they are now in "my toolbox" for future

reference and use. It is clear that my work on this competency has provided me a foundation for future learning and growth.

REFERENCES

- Angus, S. D., & Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *British Journal of Educational Technology*, 40(2), 255-272.
- Baehr, M. (2009). *Distinctions between assessment and evaluation*. Retrieved from http://matcmadison.edu/cetl/resources/archive/efgb/4/4_1_2.htm
- Baker, E. L., & Herman, J. L. (2003). A distributed evaluation model. In G. D. Haertel & B. Means (Eds.), *Evaluating educational technology: Effective research designs for improving learning*. New York, NY: Teachers College Press.
- Barbera, E. (2009). Mutual feedback in e-portfolio assessment: an approach to the netfolio system. *British Journal of Educational Technology*, 40(2), 342-357.
- Chang, C.-C., & Tseng, K.-H. (2009). Use and performances of Web-based portfolio assessment. *British Journal of Educational Technology*, 40(2), 358-370.
- Cole, P. (1992). Constructivism revisited: A search for common ground. *Educational Technology*, 32(2), 27-34.
- Cook, T. D. (2003). Why have educational evaluators chosen not to do randomized experiments? *The ANNALS of the American Academy of Political and Social Science*, 589(1), 114-149.
- Creighton, T. B. (2007). *Schools and data: the educator's guide for using data to improve decision making*. Thousand Oaks, CA: Corwin Press.
- 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.
- Draper, S. W. (2009a). Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology*, 40(2), 285-293.
- Draper, S. W. (2009b). What are learners actually regulating when given feedback? *British Journal of Educational Technology*, 40(2), 306-315.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). Boston, MA: Pearson.
- Freedman, R. L. H. (1998). *Constructivist Assessment practices*. Retrieved from <http://www.eric.ed.gov/>
- Haertel, G. D., & Means, B. (2003). *Evaluating educational technology: Effective research designs for improving learning*. New York, NY: Teachers College Press.
- Heineke, W. F., & Blasi, L. (2001). *Methods of evaluating educational technology*. Greenwich, CT: Information Age Pub.

- Heinze, A., & Heinze, B. (2009). Blended e-learning skeleton of conversation: Improving formative assessment in undergraduate dissertation supervision. *British Journal of Educational Technology*, 40(2), 294-305.
- Hense, J., Kriz, W. C., & Wolfe, J. (2009). Putting theory-oriented evaluation into practice: A logic model approach for evaluating SIMGAME. *Simulation Gaming*, 40(1), 110-133.
- Janesick, V. J. (2006). *Authentic assessment*. New York, NY: Peter Lang Publishing.
- Jonassen, D. (1991). Objectivism vs constructivism: Do we need a new philosophical paradigm? *Educational Technology, Research and Development*, 39(3), 5-13.
- Jordan, S., & Mitchell, T. (2009). e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback. *British Journal of Educational Technology*, 40(2), 371-385.
- Loddington, S., Pond, K., Wilkinson, N., & Willmot, P. (2009). A case study of the development of WebPA: An online peer-moderated marking tool. *British Journal of Educational Technology*, 2, 329-341. Retrieved from 10.1111/j.1467-8535.2008.00922.x
- Love, A. J. (1991). *Internal evaluation: building organizations from within*. New York, NY: Sage Publications.
- Marriott, P. (2009). Students' evaluation of the use of online summative assessment on an undergraduate financial accounting module. *British Journal of Educational Technology*, 40(2), 237-254.
- Matusevich, M. N. (1995). School reform: What role can technology play in a constructivist setting? Retrieved from <http://delta.cs.vt.edu/edu/fis/techcons.html>
- McNeil, K. A., Newman, I., & Kelly, F. J. (1996). *Testing research hypotheses with the general linear model*. Carbondale, IL: Southern Illinois University Press.
- McNeil, K. A., Newman, I., & Steinhauer, J. (2005). *How to be involved in program evaluation: What every administrator needs to know*. Lanham, MD: ScarecrowEducation.
- NCREL (2004). *Assessment in a constructivist classroom*. Retrieved from <http://www.ncrel.org/sdrs/areas/issues/methods/assment/as7const.htm>
- Ng'ambi, D., & Brown, I. (2009). Intended and unintended consequences of student use of an online questioning environment. *British Journal of Educational Technology*, 40(2), 316-328.
- Patton, M. Q. (1997). *Utilization-focused evaluation: the new century text*. New York, NY: Sage Publications.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

- Rumberger, R. W. (2003). The advantages of longitudinal design. In G. D. Haertel & B. Means (Eds.), *Evaluating educational technology: Effective research designs for improving learning*. New York, NY: eachers College Press.
- Shephard, K. (2009). e is for exploration: Assessing hard-to-measure learning outcomes. *British Journal of Educational Technology*, 40(2), 386-398.
- Stufflebeam, D. L., & Wingate, L. A. (2005). A self-assessment procedure for use in evaluation training. *American Journal of Evaluation*, 26(4), 544-561.
- Walvoord, B. E. (2004). *Assessment clear and simple: A practical guide for institutions, departments, and general education*. San Francisco, CA: Jossey-Bass.
- Whitelock, D. (2009). Editorial: e-assessment: developing new dialogues for the digital age. *British Journal of Educational Technology*, 40(2), 199-202.
- Williams, J. B., & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology*, 40(2), 227-236.
- Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practical guidelines* (2nd ed.). New York, NY: Longman.