4-1-2009

# Assessing the Quality of Doctoral Dissertation Literature Reviews in Instructional Technology

M. Harrison Fitt
*Utah State University*

Andrew E. Walker
*Utah State University*

Heather M. Leary
*Utah State University*

# Assessing the Quality of Doctoral Dissertation Literature Reviews in Instructional Technology

M. Harrison Fitt, Andrew E. Walker, and Heather M. Leary
*Utah State University*

*Please address all correspondence to:*
Andrew Walker
Utah State University
Department of Instructional Technology and Learning Sciences
Logan, UT 84322
Phone: 435-797-2614
Fax: 435-797-2693
Email: andy.walker@usu.edu

Abstract

*Assessment of the doctoral dissertation literature review provides insight into a student's preparation for future work as a researcher. In 2004, efforts to assess the quality of literature reviews in doctoral dissertations were pioneered by Boote & Beile. Their work represents an important response to the call for improved research skills among emerging scholars. The purpose of this study is to replicate their work in a focused area of education research, specifically Instructional Technology, and to examine the inter-rater reliability of the rubric. The findings suggest that dissertation literature reviews in Instructional Technology show the same need for improvement as dissertation literature reviews from education as a whole. Potential avenues of research are identified as well as improvements for the rubric.*

# Assessing the Quality of Doctoral Dissertation Literature Reviews in Instructional Technology

M. Harrison Fitt, Andrew E. Walker, and Heather M. Leary
*Utah State University*

The doctoral dissertation is the culminating written assessment of a PhD candidate's educational experience. As part of the doctoral dissertation, the literature review provides a unique vantage point to examine the overall quality of a student's preparation for future work as an independent researcher. It is an indicator of their ability to critically analyze their research area, seek out new relationships between seemingly unconnected phenomena, resolve ambiguities, frame their own research and pose timely research questions. While little work has been done on other components of a dissertation, there is small but growing body of literature emerging on the assessment of the doctoral literature review (see Hart, 1998; Boote & Beile, 2004, 2005; Holbrook, Bourke, Fairbairn, & Lovat, 2007; Fitt, Bentley, & Gardner, 2008) as well as students' perception, experiences with, and understanding of dissertation literature reviews (see Bruce, 1994, 2001; Nelson, 2007).

Boote & Beile's (2004, 2005) important research about the quality of literature reviews in doctoral dissertations utilizes the *Literature Review Scoring Rubric,* which they created. Initially used to evaluate literature reviews of doctoral dissertations in education as a broad field, it is now being applied in areas of inquiry such as nursing (Bowman, 2007), music (Freer & Barker, 2008), information systems (Levy & Ellis, 2006), and the teacher professional continuum (Stuessy, 2007). It has also been used to guide the development of a survey for graduate students' perceptions of their preparation for conducting literature review for their thesis or dissertation (Nelson, 2007). However, Boote & Beile applied their rubric to dissertations from the field of education, and only chapter 2 of the dissertation, the chapter which traditionally contains the literature review. Additionally, their stratified random sampling process yielded only quantitative dissertations. It remains to be seen if the rubric works well for specific fields of inquiry within education and qualitative or mixed-methods designs, and if examining the entire dissertation as opposed to the literature review alone impacts score. Finally, while the rubric has been used, its reliability has not been assessed when used by evaluators other than the original authors of the rubric.

To that end, the purpose of this work is to extend the work of Boote & Beile by (a) applying the *Literature Review Scoring Rubric* to a more narrow field of inquiry within education research, (b) evaluating each dissertation based on chapter 2 alone and the entire work, (c) examining dissertations that incorporate a range of methodologies, specifically qualitative, mixed-methods, and quantitative (d) analyzing the inter-rater reliability of the Boote & Beile rubric.

## Literature Review

Several initiatives, such as the Carnegie Initiative on the Doctorate and the Pew Charitable Trust, have led to an increased awareness of the need for doctoral students to develop

more sophisticated research skills (Golde & Walker, 2006; Walker, Golde, Jones, Bueschel, & Hutchings, 2008). More recently, the American Educational Research Association (AERA) and the National Academy of Education (NAEd) have undertaken a large-scale study examining education research doctorate programs in the United States to aide the understanding of the "substance and quality of education research doctorate programs" (AERA, 2008). Also, as editors of *Review of Educational Research*, LeCompte and colleagues assert that current and comprehensive literature reviews can make a valuable contribution to areas of inquiry (LeCompte, Klingner, Campbell, & Menk, 2003) and specifically encourage emerging scholars to learn the necessary skills to write such works. Several authors parallel LeCompte et al, adding suggestions for ways to improve literature reviews (e.g., Alton-Lee, 1998; Cooper, 1982; Cooper, 1985; Hart, 1999; Lather, 1999; Lester, 2002; Light & Pilmer, 1982; Locke, Spirduso, & Silverman, 2007; Strike & Posner, 1983). All of this work underscores the widespread interest in improving the quality of literature reviews, much of which is focused on doctoral students and emerging scholars.

Practically speaking, if a researcher is unable to identify what work has already been done in the field and what avenues of scholarly inquest have yet to be investigated, there is a diminished capacity for the researcher to produce useful and timely research (Alton-Lee, 1998; LeCompte, Klingner, Campbell, & Menk, 2003). In particular, new doctoral recipients who have not mastered the skills of reviewing and synthesizing current literature run the risk of not understanding the most pressing issues within the field (Lather, 1999; LeCompte et al., 2003).

*The Dissertation Literature Review*

The doctoral dissertation is a singular opportunity for a PhD candidate to demonstrate they have the capabilities and necessary preparation for independent scholarly work (Isaac, Quinlan, & Walker, 1992; see also Association of American Universities, 1998 and Council of Graduate Schools, 1997, 2004). As part of the only tangible evidence of a candidate's research (Bruce, 1994; Hart, 1999), the literature review allows them to showcase their ability to critically analyze what work has already been done in the field and how it was conducted, what lines of inquiry have yet to be investigated, their ability to synthesize research from their specific field as well as others, and their ability to resolve ambiguities in the vocabulary and literature (Creswell, 2008; Fraenkel & Wallen, 2006; Gall, Gall, & Borg, 2007; Gay, Mills, Airasian, 2006; Johnson & Christensen, 2004; McMillan,2008; Schumacher & McMillan, 2006). As such, the doctoral dissertation literature review can be viewed as one barometer of the overall health of doctoral research training.

While the stylized literary structure of the dissertation literature review (Lovat, 2004) may vary dependent of the type of research the student is involved in (i.e., qualitative, quantitative, or mixed methods), all forms serve similar functions. Primarily, it lays the foundation and provides a context to the research question posed in the dissertation. Additionally, the dissertation literature review allows the candidate to display the high levels of critical thinking skills and sophisticated reasoning required to be a successful researcher for scrutiny and assessment (Isaac et al, 1992; Hart, 1998).

*Rubrics as Assessment Tools*

As a self-reflexive assessment tool, a well-constructed rubric can help students develop independent critical thinking (Andrade, 2000; Arter, 1993; Huba & Freed, 2000, Simon & Forgette-Giroux, 2001; Stevens & Levi, 2005), a skill necessary to crafting a well-written dissertation literature review. The same rubric can also aid educators in the application of consistent assessment standards and can help clarify potential areas for improvement (Lovitts, 2006/2007; Mertler, 2001; Moskal, 2000; Moskal & Leydens, 2000; Simon & Forgette-Giroux, 2001; Stevens & Levi, 2005, Tierney & Simon, 2004). The 12-item *Literature Review Scoring Rubric* (Boote & Beile, 2004, 2005) is an important means of improving literature reviews in doctoral dissertations as it represents a way for faculty supervisors to clarify their expectations to the doctoral candidate (Lovitts, 2006/2007; Simon & Forgette-Giroux, 2001). The rubric can guide candidates in understanding the process of conducting and writing the literature review. Finally, it can serve as an important educational tool for the candidate to refer to when asked to self-assess their own work and how they might improve on it (Moskal & Leydens, 2000; Simon & Forgette-Giroux, 2001; Steven & Levi, 2005).

Guided by the common call for improved research skills in education, the *Literature Review Scoring Rubric* (Boote & Beile, 2004, 2005) is an adaptation of Christopher Hart's (1999) important work in which he outlines at least eleven of the distinct purposes of a literature review in a thesis or dissertation. Boote & Beile (2004, 2005) expanded the purposes to a list of twelve criteria and divided them into five categories: Coverage, Synthesis, Methodology, Significance, and Rhetoric. (Note that categories to not have the same number of criteria.) The 12 criteria on the *Literature Review Scoring Rubric* are mostly scored on a scale of one (low) to three (high), see Appendix A for details. An exception to this, Criterion H, is scored on a four-point scale.

In Part A of their original study, Boote & Beile (2004) examined 30 dissertations from a stratified random sample and conducted a citation analysis for each. For Part B, a purposeful sample of 12 dissertations was then selected from the 30 original dissertations (four from each of the three universities represented in the stratified random sample). These 12 dissertations were then evaluated using the *Literature Review Scoring Rubric* (Boote & Beile, 2004, 2005). In regards to Part B, the results of their study reveal that while there was a wide range of quality scores, there was a common failure to synthesize, critique, or explain relevant literature and methodologies. While Boote & Beile's *Literature Review Scoring Rubric* is a vital part of emerging research, the original study has some limitations in that it examined only dissertations using quantitative methods, the *n* was small, and the reliability of the rubric has not been fully established.

Researchers in Instructional Technology come from and draw on many disciplines including but not limited to computer science, artificial intelligence, technical writing, psychology, and education. The cross-disciplinary nature of Instructional Technology makes it a natural bridge between the general field of education research and more focused areas of inquiry within education research and beyond. The consequences of scholars who do not possess sound review skills can be far reaching, especially in an interdisciplinary field such as Instructional

Technology. Indeed, if sophisticated literature review skills are important for the field of education in general, they are magnified for Instructional Technology due to the inherent danger of parallel effort in a cross-disciplinary field. To further the understanding of the quality of literature reviews in education doctoral dissertations, we conducted an instrument design study replicating Part B of Boote & Beile's 2004 study. Following are the research questions for which we sought answers:

1. What differences exist between dissertations from Instructional Technology and education as a general field of study as measured by Boote & Beile?
2. What differences exist between the scores derived from chapter 2 alone and scores derived from the entire dissertation?
3. What score differences exist among literature reviews using quantitative, qualitative, or mixed-method study designs?
4. Do scores change differently when derived from chapter 2 alone or the entire dissertation based on the methodology employed?
5. What is the inter-rater reliability of the Boote & Beile rubric?

**Methods**

This research is an instrument design study of the *Literature Review Scoring Rubric* (Boote & Beile, 2004/2005) for the assessment of doctoral dissertation literature reviews. The specific contribution includes a focus on the field of Instructional Technology. Using *Dissertation Abstracts*, a list of 333 dissertations from Instructional Technology awarded during the years 2006 and 2007 was compiled. From those, 30 dissertations were randomly selected for evaluation. The lead researcher then removed any identifying information so "blind" evaluations could be conducted.

A team of five reviewers consisting of four doctoral students and one faculty member in Instructional Technology were trained on one of the dissertations using the *Literature Review Scoring Rubric* (Boote & Beile, 2004, 2005). To do this, each reviewer evaluated the first dissertation using the rubric and then met to discuss the nuances of applying the rubric and come to a consensus for the score. From there, each dissertation was scored by two total raters. After a "first-pass" scoring, pairs of evaluators discussed each dissertation until they reached consensus. With consistent discrepancies in first-pass scores, due to missed data rather than direct disagreement, a decision was made to always have pairs of raters and always discuss results until consensus was achieved (Yancey, 1999; Stemler, 2004). Dissertations were then given an overall score from 12-37 points by adding values across all 12 criteria. In both the first and second pass scoring, data were reported twice for each rater on each dissertation. One score followed Boote & Beile's initial work (2004), examining only the second chapter or literature review. The other score was drawn from any portion of the dissertation, most frequently the introduction and methods. Upon completion of the evaluations, descriptive statistics were computed, overall rubric scores were analyzed using a factorial ANOVA, and inter-rater reliability was computed using an intra-class correlation. Also, Boote & Beile provided their raw scores from the 12 dissertations in their original study so comparisons could be made between research findings.

# Results

Effect sizes are reported as Cohen's *d* using the pooled estimate of the population standard deviation as the denominator. The alpha level for statistical significance tests was set at .05. Of the 30 dissertations, three were dropped. One dissertation was dropped because it was used for training the raters and the other two were dropped for methodological reasons. One of the methodological drops employed a meta-analysis, which as a research form, inherently aligns with prescriptions from the rubric. As an example, a meta-analyses requires full disclosure of inclusion and exclusion decisions (Cooper & Hedges, 1994) which aligns directly with criterion A. The other was a discourse analysis which contained no identifiable literature review chapter making it impossible to score using the *Literature Review Scoring Rubric* (Boote & Beile, 2004, 2005). Of the remaining 27 dissertations, six were quantitative in design, 12 were qualitative, and nine were mixed methods.

In regards to research question one the mean score for the Instructional Technology (*N*=27) dissertations was 19.96 (*SD* =3.16) was substantially lower than the mean for educational dissertations as a whole (*N*=12) from Boote & Beile 24.08 (*SD*=6.05) (Boote & Beile, 2004). Note there are differences in both the means and the standard deviation. With respect to the standard deviation some of this may be due to positive bias, a result of a much smaller sample size for the prior data. However, that is not the case with the mean. Placing these differences in perspective, there is what Cohen (1988) described as a "large" effect size favoring Boote & Beile's data (*d*=0.97). This may well look like an indictment against Instructional Technology dissertations, but additional discussion is warranted. There are three potential causes of score differences: (1) the field of Instructional Technology, which seems unlikely given the high scores for Instructional Technology dissertations coded by Boote & Beile, (2) different selection criteria for the studies with no attempt to stratify a range of program quality or (3) the use of different raters.

To address research questions two and three, rubric scores were analyzed using a 3x2 factorial ANOVA with the factors being research design (quantitative, qualitative, or mixed method) and coverage (chapter two only, all chapters). Mean statistics are reported in Table 1. The results included a significant main effect for coverage $F(1, 53)=7.01$, p=0.011, with dissertations obviously scoring higher with all chapters analyzed (*M*=22.07) than the literature review chapter alone (*M*=19.96). There was also a main effect for research design $F(2, 52)=4.48$, p=.017, in which both quantitative (*M*=22.08) and qualitative (*M*=21.71) work outscored mixed methods (*M*=19.36) approaches.
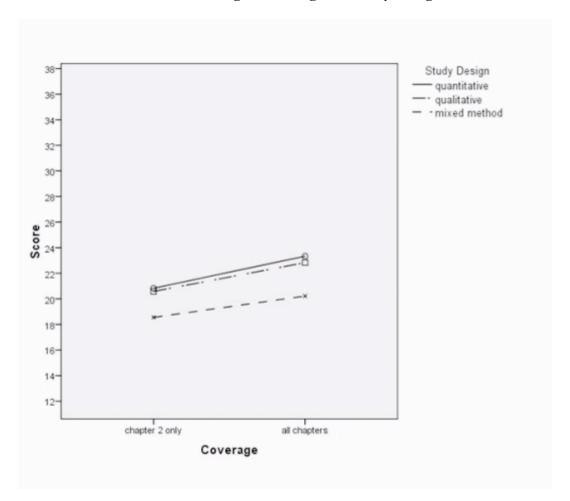
TABLE 1. *Descriptive Statistics of Consensus Dissertation Scores*

| | Coverage (0-37 points) | |
|---|---|---|
| Dissertation Style | Chapter two only *m*(*sd*) | All chapters *m*(*sd*) |
| Quantitative (*n*=12) | 20.83(2.93) | 23.33(2.25) |
| Qualitative (*n*=6) | 20.58(3.18) | 22.83(2.92) |
| Mixed methods (*n*=9) | 18.56(3.13) | 20.22(2.22) |

In response to research question two and three, the main effects suggest that there is a difference based on both dissertation coverage and methods employed. However, placing this main effect in context is important. The largest pairwise difference favors quantitative dissertations over mixed methods dissertations when all chapters are included. While the effect size comparison between these two is quite large ($d$=1.39) and constitutes a ten percent increase in points, neither mean is very flattering out of the possible 37 on the rubric. In short, the dissertation scores are poor and variations may be statistically significant, but in terms of practical significance, the scores are still poor.

As can be seen in Figure 1, there was no interaction effect for coverage x study design $F(2, 52)$=0.90, p=0.92. The clear response to research question 4 is no, the changes in methodology scores at each coverage type were almost parallel for quantitative, qualitative and mixed-method designs. Although caution should be exercised as we focused on Instructional Technology work alone, it is possible that dissertations in the broader field of education will score similarly with various research designs irrespective of coverage.

FIGURE 1. *Dissertation Scores according to Coverage and Study Design*

To examine inter-rater reliability and address research question five, an intra-class correlation was used. Because all elements of the rubric are not all on the same scale (one item is scored from 1 to 4 whereas the remaining are all 1 to 3) the analysis was run on the total scores for each dissertation. The resulting intra-class correlation on inter-rater reliability was not at all flattering (.344) and indicates there was very little agreement on first pass scoring of these dissertations. Possible reasons for this are discussed below.

## Conclusions

The results indicate little difference exists between the scores for chapter 2 and the overall dissertation when examined as a whole, providing some partial vindication for Boote & Beile's (2004) focus on chapter 2 in their initial work. Additionally, this may help greatly with use of the rubric since reading through the entire document, even when focused on elements of the rubric, can be incredibly time consuming. The low inter-rater reliability of the first-pass scores was likely due to a combination of factors. First, reviewers were not as familiar with the rubric as its creators. Second, as emerging scholars they may have had less consensus about interpreting elements of the rubric as compared with those who scored dissertations in Boote & Beile's (2004) study.

Finally, the rubric itself may have some inherent shortcomings. The difference between scale levels is not conceptually similar for the varying dimensions. For example, to earn a score of two on many of the criterion, a student must "discuss" the criterion. However, for criterion E, "acquired and enhanced the subject vocabulary," discussing the criterion earns a three. Conversely, to earn the higher score of three for most criterion, a student must critique or "critically examine" the criterion. The exception to this is criterion G, "synthesized and gained a new perspective on the literature," which requires a student to critique the literature for a score of two (instead of three). Particularly problematic is the scale level between a score of two and three for criterion D, "placed the research in the historical context." To earn a two, the student need merely "mention" the history of the topic, while the student must make a large cognitive leap to earn a three as it requires them to "critically examine" the history of the topic. Even more troubling, one of the 12 criterion is scored on a scale of four instead of three rendering more sophisticated statistical analysis of the data impractical.

The unequal conceptual differences in scale levels for various criteria makes it difficult in some instances for raters to determine the appropriate score to give the dissertation. In order to improve the reliability of the rubric, clarification of the scale levels is needed, perhaps creating four scale levels for all of the criteria, allowing for a finer gradation of acceptable versus exemplary work. This would also allow for individual scores to be compared in a more meaningful manner. It is quite possible, however, that a high intra-class correlation score is out of reach due to the inherent complexity of the task. In parallel work, it is rare for meta-analysts to use a single rater. Using multiple coders and then reaching consensus may just be the required norm.

The low overall scores may point to a systemic issue within the process of doctoral education in regards to how students are being taught literature review skills. Either the students

are not learning them or faculty are not teaching them. However, instead of being a sweeping indictment of the current system, these findings offer the opportunity for a closer examination of our current practices in educating future scholars and researchers. Indeed, if Phelps (2007) is correct, an increased focus on "the lost art of the literature review" can help set education research back on its rightful course.

The lack of an interaction effect between coverage and methodology is puzzling. While the process of engaging in a literature review for quantitative, qualitative, and mixed-methods designs is similar the write-up has potential for great differences. Given the emergent nature of qualitative research, there should be more indication of literature review elements in the methods, analysis, results, and conclusion sections. It is unclear why score differences for coverage remained parallel for these Instructional Technology dissertations.

This replication study has several limitations. The factorial ANOVA doesn't meet the requirement of equal cell sizes, which may have impacted the statistical significance of the main and interaction effects. While the focus on Instructional Technology is a good first step caution should be used when generalizing to education as a whole or other focused areas of inquiry. Differences from the Boote & Beile study could be due to the fact that different raters were used, not because a different content area was examined. Additionally, this study did not deal with the relevance of the literature review in relation to the dissertation research and the possibility of a literature review scoring well on the rubric but lacking relevance to the methods, data collection results and conclusion, a criticism raised by Maxwell (2006). Anecdotally, incongruence between the literature review and other portions of the dissertation was something reviewers observed even when they were not prompted to do so. Informally, the raters noted the detailed use of the Boote & Beile rubric was a valuable learning experience in terms of crafting their own dissertations, a phenomenon which is supported by research about the benefits of using rubrics in assessment (Moskal & Leydens, 2000; Stevens & Levi, 2005). They agreed its use in the research study improved their own awareness of key elements of a quality doctoral dissertation literature review.

Few would question the statement that well-written, sophisticated literature reviews lead to good research (Alton-Lee, 1998; Lather, 1999; LeCompte, Klingner, Campbell, & Menk, 2003). The *Literature Review Scoring Rubric* (Boote & Beile, 2004, 2005) represents a means for assessing the quality of the dissertation literature review. However, much more work needs to be done to establish the reliability and validity of the rubric. In light of the research results the authors propose that future scholarly research should focus on the following:

- Identifying the reasons students were able to score well on the rubric, such as the mentoring model and process of their dissertation chairs
- Establishing the predictive validity of the rubric by uncovering whether or not the study design and magnitude of measured outcomes is related to the quality of the literature review
- Examining the construct validity of the rubric to ensure all of the important facets of a dissertation literature review are represented in the rubric
- Determining if the rubric is useful in assessing non-traditional formats such as multiple-paper dissertations or design-based research

- Revising the rubric so that the scale levels are conceptually consistent and conducting further reliability and validity studies on the revised rubric

As a response to the call for increased scholarship in education research through more careful attention the to literature reviews, the *Literature Review Scoring Rubric* (Boote & Beile, 2004, 2005) has proven useful for evaluating dissertation literature reviews, at least within the field of Instructional Technology, with respect to a diverse set of research methods and even when used with the literature review chapter alone. This work raises several questions of its own, most notably why qualitative and mixed-methods designs fail to show more dramatic improvement when the entire dissertation is scored. In addition to posing new questions, this research also represents an important step forward in the valid and reliable measurement of literature review efforts by students.

**REFERENCE LIST**

Alton-Lee, A. (1998). A troubleshooter's checklist for prospective authors derived from reviewers' critical feedback. *Teaching and Teacher Education*, *14*(8), 4.

American Educational Research Association (2008). *AERA, NAEd Launch Assessment of Education Research Doctorate Programs with National Science Foundation Support.* Press release. Retrieved on March 3, 2009, from http://aera.net/newsmedia/Default.aspx?menu_id=60&id=6192

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 13-18.

Arter, J. (1993). *Designing scoring rubrics for performance assessments: The heart of the matter*. Portland, OR: Northwest Regional Education Laboratory. (ERIC Document Reproduction Service No. ED 358 143)

Association of American Universities. (1998). *Committee on graduate education report and recommendations.* Washington, DC: Author.

Boote, D. & Beile, P. (2004, April). *The quality of dissertation literature reviews: A missing link in research preparation*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Boote, D. & Beile, P. (2005). Scholars Before Researchers: On the Centrality of the Dissertation Literature Review in Research Preparation. *Educational Researcher*, 34(6), 3-15.

Bowman, K. G. (2007). A research synthesis overview. *Nursing Science Quarterly,* 20, 171-176.

Bruce, C. S. (1994). Research students' early experiences of the dissertation literature review. *Studies in Higher Education, 19*(2), 217-229.

Bruce, C. S. (2001). Interpreting the scope of their literature reviews: Significant differences in research students' concerns. *New World Library, 102*(1163/1164), 158-165.

Carnegie Commission on Higher Education (2001). *A Classification of Institutions of Higher Education,* 2000 Ed. Menlo Park, CA.

Carnegie Foundation (2003). *Carnegie Initiative on the Doctorate*. Retrieved March 15, 2008, from http://www.carnegiefoundation.org/programs/index.asp?key=29

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cooper, H., & Hedges, L. (1994). *The Handbook of research synthesis*. New York: Russell Sage Foundation.

Cooper, H. M. (1982). Scientific Guidelines for Conducting Integrative Research Reviews. *Review of Educational Research*, *52*(2), 291-302.

Cooper, H. M. (1985). *A taxonomy of literature reviews*. Annual meeting of the American Educational Research Association. Chicago, ERIC Document Reproduction Services No. ED254541.

Cooper, H. M. (1988) .The structure of knowledge synthesis. *Knowledge in Society, 1*, 104-126.

Cooper, H. M. (1989). *Integrating Research: a guide for literature reviews*, *2nd Ed*. Newbury Park, CA: Sage.

Council of Graduate Schools (1997). *The role and nature of the doctoral dissertation: A policy statement.* Washington, DC: Author.

Council of Graduate Schools (2004). *The doctor of philosophy degree: A policy statement.* Washington, DC: Author.

Creswell, J. (2008). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (3rd ed.). New Jersey: Pearson

Fitt, M. H., Bentley, J. P., & Gardner, J. (2008). Rays of change: Towards a better framework for doctoral dissertation literature reviews in Instructional Technology. *Proceedings of the Conference of the Association for Educational Communications and Technology*, 1.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to Design and Evaluate Research in Education* (6th ed.) New York: McGraw Hill.

Freer, P. K. & Barker, A. (2008). An instructional approach for improving the writing of literature reviews. *Journal of Music Teacher Education*, *17*(2), 69-82.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational Research: An Introduction* (8th ed.). New York: Pearson.

Gay, L. R., Mills, G., & Airasian, P. W. (2006). *Educational Research* (8th ed.). New York: Pearson.

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, *5*(10), 6.

Glass, G. V. (1977). Integrating Findings: The meta-analysis of research. *Review of Research in Education*, *5*, 29.

Golde, C. M., Jones, L., Conklin Bueschel, A., & Walker, G. E. (2006). The challenges of doctoral program assessment: Lessons from the Carnegie Initiative on the Doctorate. In P. L. Maki & N. A. Borkowski (Eds.), *The assessment of doctoral education: Emerging criteria and new models for improving outcomes,* (pp. 53-82). Sterling, VA: Stylus.

Golde, C. M., & Walker, G. E. & Associates (2006). *Envisioning the future of doctoral education: Preparing stewards of the discipline. Carnegie essays on the doctorate.* San Francisco: Jossey-Bass.

Hart, C. (1999). *Doing a literature review: Releasing the social science research imagination.* London: SAGE Publications.

Holbrook, A., Bourke, S., Lovat, T. & Dally, K. (2004). Qualities and characteristics in the written reports of doctoral thesis examiners. *Australian Journal of Educational and Developmental Psychology, 4,* 126-145.

Holbrook, A., Bourke, S., Fairbairn, H., & Lovat, T. (2007). Examiner comment on the literature review in Ph.D. theses. *Studies in Higher Education, 32,* 337-356.

Huba, M.E., & Freed, J. E. (2000). *Learner-centered assessment on college campuses: Shifting the focus from teaching to learning.* Boston: Allyn & Bacon.

Isaac, P. D., Quinlan, S. V., & Walker, M. M. (1992). Faculty perceptions of the doctoral dissertation. *Journal of Higher Education, 63*(3), 241-268.

Johnson, B., & Christensen, L. (2004). *Educational Research: Quantitative, Qualitative, and Mixed Approaches* (2nd ed.). New York: Pearson.

Lather, P. (1999). To be of use: The work of reviewing. *Review of Educational Research*, *69*(1), 6.

LeCompte, M. D., Klingner, J. K., Campbell, S. A., & Menk, D. W. (2003). Editor's Introduction. *Review of Educational Research*, *73*(2), 2.

Lester, J. D. (2002). *The essential guide: Research writing across the disciplines.* New York: Longman.

Levy, Y. & Ellis, T. J. (2006). *Towards a framework of literature review process in support of information systems research.* Paper presented at the 2006 Informing Science and IT Education Joint Conference.

Libutti, P., & Kopola, M. (1995). The doctoral student, the dissertation, and the library: A review of the literature. *Reference Librarian, 48*, 5-25.

Light, R. J., & Pillemer, D. (1982). Numbers and narrative: Combining their strengths in research reviews. *Harvard Educational Review, 52*, 1-26.

Locke, L. F., Spirduso, W. W., & Silverman, S. J. (2007). *Proposals that work: A guide for planning dissertation and grant proposals* (5th ed.). New Jersey: SAGE Publications.

Lovat, T. (2004). 'Ways of Knowing' in doctoral examination: How examiners position themselves in relation to the doctoral candidate. *Australian Journal of Educational and Developmental Psychology, 4,* 146-152.

Lovitts, B. E. (2006). Making the implicit explicit: Faculty's performance expectations for the dissertation. In P. L. Maki & N. A. Borkowski (Eds.), *The assessment of doctoral education: Emerging criteria and new models for improving outcomes,* (pp. 163-196). Sterling, VA: Stylus.

Lovitts, B. E. (2007). *Making the implicit explicit: Creating performance expectations for the dissertation*. Sterling VA: Stylus.

Maxwell, J. A. (2006). Literature reviews of, and for, educational research: A commentary on Boote and Beile's "Scholars Before Researchers." *Educational Researcher, 35*(9), 28-31.

McMillan, J. (2008). *Educational Research: Fundamentals for the Consumer* (5th ed.) Boston: Pearson.

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research and Evaluation, 7*(25).

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research and Evaluation, 7*(10). Retrieved March 3, 2009, from http://pareonline.net/getvn.asp?v=7&n=10

Moskal, B. M. (2000). Scoring rubrics: What, when, & how? *Practical Assessment: Research & Evaluation 7*(3).

Nelson, F. F. (2007). *Perceptions of graduate students regarding the challenges of conducting research in higher education*. Unpublished thesis. Southern Illinois University at Carbondale.

Pan, M. L. (2003). *Preparing literature reviews: Qualitative and quantitative approaches.* Los Angeles: Pyrczak.

Pew Charitable Trust (2001). *Re-envisioning the Ph.D.* Retrieved March 15, 2008 from http://www.grad.washington.edu/envision/

Phelps, R.P. (2007). The dissolution of education knowledge. *Educational Horizons, 85*(4), 232-247.

Seels, B. B., & Richey, R. C. (1994). *Instructional technology: The definition and domains of the field.* Washington, DC: AECT.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5-11.

Slavin, R. E. (1995). Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology, 48*(1), 9-18.

Simon, M.& Forgette-Giroux, R. (2001). A rubric for scoring postsecondary academic skills. *Practical Assessment, Research & Evaluation, 7*(18). Retrieved March 5, 2009 from http://PAREonline.net/getvn.asp?v=7&n=18

Stuessy, C. L. (2007). Literature Review as an Invitation to Inquiry: Beginning Research on Teacher Professional Continuum. *School Science and Mathematics, 107*(2), 42.

Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: Focusing on consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation, 9*(2), Retreived February 28, 2009 from http://PAREonline.net/getvn.asp?v=9&n=2

Schumacher, S., & McMillan, J. (2006). *Research in Education: Evidence Based Inquiry.* Boston: Pearson.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation, 9*(4). Retrieved February 28, 2009 from http://PAREonline.net/getvn.asp?v=9&n=4

Stevens, D. D., & Levi, A. J. (2005). *Introduction to Rubrics*. Sterling, VA: Stylus.

Strike, K., & Posner, G. (1983). Types of synthesis and their criteria. In S. Ward & L. Reed (Eds.), *Knowledge structure and use: Implications for synthesis and interpretation* (pp. 343-362). Philadelphia: Temple University Press.

Walker, G. E., Golde, C. M., Jones, L., Bueschel, A. C., & Hutchings, P. (2008). *The formation of scholars: Rethinking doctoral education for the twenty-first century*. San Francisco, CA: Josey-Bass.

Yancey, K.B. (1999). "Looking back as we look forward: Historicizing writing assessment." *College Composition and Communication, 50,* 483-503.