



2012-06-28

Evaluation in Computer-Assisted Language Learning

Benjamin L. McMurry
Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Educational Psychology Commons](#)

BYU ScholarsArchive Citation

McMurry, Benjamin L., "Evaluation in Computer-Assisted Language Learning" (2012). *All Theses and Dissertations*. 3306.
<https://scholarsarchive.byu.edu/etd/3306>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Evaluation in Computer-Assisted Language Learning

Benjamin L. McMurry

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Peter J. Rich, Chair
Richard Edward West
Neil J. Anderson
K. James Hartshorn
David Dwayne Williams

Department of Instructional Psychology and Technology

Brigham Young University

August 2012

Copyright © 2012 Benjamin L. McMurry

All Rights Reserved

ABSTRACT

Evaluation in Computer-Assisted Language Learning

Benjamin L. McMurry

Department of Instruction Psychology and Technology, BYU

Doctor of Philosophy

Evaluation of Computer-assisted language learning (CALL) needs to be scrutinized according to the same standards of evaluation as other professional materials. Evaluation can be divided into two distinct, yet similar, categories: formal (following a prescribed evaluation model) and informal. The aim of this dissertation is two-fold. The first purpose is to benefit the field of CALL by situating CALL evaluation in the context of frameworks used formal evaluation. The second purpose is to discover informal evaluation practices of CALL practitioners. First, with regard to formal evaluation of CALL materials, practices and insights from the field of evaluation would help CALL researchers and practitioners to conduct systematic evaluations that report findings that other researchers and practitioners find useful. An evaluation framework is proposed that includes common evaluation tasks conducted by evaluators in the field of formal evaluation to produce a workflow model for designing and conducting evaluations in CALL. Second, regarding the informal evaluation of CALL materials, learning about the processes of teachers when evaluating CALL for classroom use will help direct developers of CALL materials, address user concerns, and may indirectly increase the quality of CALL materials.

After looking at this two-fold question—formal and informal evaluation of CALL materials—we found that formal evaluation in CALL may benefit from the adoption of evaluation practices from formal evaluation literature. Regarding informal evaluation, we found that teachers consider pedagogy, accessibility, and authenticity when reviewing CALL resources and activities for consideration for use in the classroom. Based on this finding we provide implications for language program administrators, teacher trainers, CALL software developers, and language teachers.

Keywords: computer-assisted language learning, technology, computer-assisted instruction, evaluation, framework, second language learning

ACKNOWLEDGEMENTS

I would like to thank my committee for their invaluable support and encouragement. They are my mentors and friends who have helped me to become a professional in second language learning and computer-assisted language learning.

Most importantly, I owe an enormous debt of gratitude to my wonderful wife who always provided encouragement and helped me stay motivated throughout my schooling. She also spent time helping locate previous research and editing my paper. Thank you!

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
Description of Structure and Content	viii
Format	viii
Background	viii
Overview of the Articles	x
ARTICLE 1: A Systematic Evaluation Framework for CALL	1
Article Abstract	2
Introduction	3
Methodology	5
Evaluation Defined and Described	6
Analysis of Popular CALL Frameworks	7
Hubbard’s Framework	8
Chappelle’s Framework	11
Framework for Designing and Conducting CALL Evaluations	13
Comparison of CALL Evaluation Frameworks to Formal Evaluation Tasks	14
Evaluands	15
Stakeholders	15
Criteria	17
Purposes	17
Evaluation Types	17
Evaluation Questions	18
Collection and Analysis of Data	18
Report of Findings	18
Evaluation Tasks	19
Identifying the Evaluand	20
Identifying Stakeholders	21
Determining the Purpose of the Evaluation	23
Selecting the Type of Evaluation	24
Setting Evaluation Criteria	29
Developing Evaluation Questions	31
Collecting and Analyzing Data	32
Reporting Evaluation Results	35
Evaluating the Evaluation	36
Discussion	37
Implications	37
Suggestions for Future Research	38
Conclusion	39
Article References	41
ARTICLE 2: Criteria Language Teachers Use When Selecting CALL Technologies	46
Article Abstract	47

Introduction.....	48
Literature Review.....	49
Teacher Education	49
Teacher Attitudes toward Technology.....	49
Teacher Experience with Technology	50
Teacher Confidence in Technology Use.....	51
Teacher Perception of Computer Use in the Classroom.....	51
Method	53
Design.....	53
Research Site.....	54
Participants.....	54
Subjectivity Statement	56
Procedures.....	57
Data Analysis.....	58
Vignettes	59
Justin.....	59
John.....	61
Shada.....	63
Hsiu Ting	64
Bettina.....	65
Jessica	67
Findings	68
Consideration for Pedagogy.....	68
Consideration of Convenience or Accessibility.....	70
Consideration of Purpose and Authenticity	72
Discussion.....	73
Limitations	73
Implications	74
Suggestions for Future Research	80
Article References.....	82
Overall Discussion.....	85
Trends in CALL Literature	85
Conclusion	86
Dissertation References	88

LIST OF TABLES

Table 1. Chapelle's (2001, 2007) Evaluands, Questions, and Evaluation types	12
Table 2. Formal Evaluation Tasks, Hubbard's (2011) Framework, and Chapelle's (2001, 2007) Framework.....	16
Table 3. Selection of Formal Evaluation Types, Descriptions, and Possible CALL Applications	27

LIST OF FIGURES

Figure 1. Hubbard's (2011) Evaluation framework.....	10
Figure 2. Framework for systematic CALL evaluation.....	14
Figure 3. Evaluation tasks with the development of evaluation questions.	31

DESCRIPTION OF STRUCTURE AND CONTENT

Format

This dissertation is in an alternate article format. In the introductory section I provide an overview of the content of and the articles. Following the introduction are the two articles each with its own figures, tables, and references. Concluding the dissertation is a final section that reviews and summarizes the two articles with a list of references at the end pertaining to the works cited in the introductory and concluding sections.

Background

Computer-assisted language learning (CALL) is the convergence of two fields: Computer-assisted instruction (CAI) and second language learning (SLL). CALL has been a buzzword for the past two decades. However, the history of CALL begins much earlier in the 1960s and 1970s. Time-shared, Interactive, Computer-Controlled Educational Television or TICCET and later TICCIT (*educational* was replaced by *information*) stands as one of the most recognized pioneer software applications in language instruction. Brigham Young University (BYU) was among the leading institutions to develop and institutionalize TICCIT with over two thousand users each semester in 1994 when it was replaced by Computerized Language Instruction and Practice Software (CLIPS).

The early 1990s saw a direct benefit of the development of personal computers and their shrinking costs. Many fields both within and outside academia began to benefit from computers, and CALL was not exempt from the phenomenon. By the turn of the millennium, computers were not only affordable and available but the internet provided access to a plethora of resources and possible interactions. Today computers and the internet are so common that they are found on mobile phones and various other devices that are small enough to fit in pockets and exponentially more powerful than their predecessors.

However, using the term CALL may now be less popular as the distinction between *traditional* materials and CALL materials fades. There is no longer a debate about whether CALL is effective or not, as CALL is not a passing fad, and CALL materials are becoming the *traditional* materials. Egbert (2005) argued that we no longer research to learn if CALL materials are more effective than traditional materials. Instead, research focuses on which practices and materials most effectively foster language learning.

Chapelle (2010) pointed out that despite the importance of high quality materials, the literature on materials evaluation in second language acquisition (SLA) was sparse. Understanding the criteria SLA teachers use in selecting, using, and evaluating CALL materials may help inform the design of such materials and their implementation by teachers, students, and administrators. As the field of CALL transitions from a passing fad to mainstream, its application needs to be scrutinized according to the same standards of evaluation as other professional materials.

While practitioners follow similar evaluation processes for both CALL materials and *non-CALL* materials, published reviews and formal evaluations of CALL still fall short of the standards upheld by mainstream evaluation practices. Though readily absent in CALL literature, questions, standards, and guidelines like those provided by ESL materials evaluation specialists such as Tomlinson (2003, 2007) are equally applicable to CALL. Even further removed are the practices and standards provided in formal evaluation literature and the Joint Committee on Standards for Educational Evaluation (JCSEE). This is not to say that CALL is unique or that it does not require special allowances when being evaluated. However, these established principles and frameworks allow for the idiosyncrasies of various fields including CALL. Their application

in CALL may serve to better illustrate the ways in which CALL materials are used and how they can be improved.

Evaluation can be divided into two distinct, yet similar, categories: formal and informal. Formal evaluation involves following established guidelines and standards. For example, the JCSEE has published standards to help evaluators conduct and report evaluations that are useful, feasible, proper, and accurate. Following a prescribed model for conducting and reporting an evaluation is a common attribute of formal evaluations and is often considered a delineating feature between it and informal evaluations. Informal evaluations happen all the time. For example, when choosing which car to purchase or which restaurant to eat at, we make value judgments. The processes are similar to those of formal evaluation though no prescribed process is required. Written reports with details explaining the process are not typically written.

The first aim of this dissertation is therefore to outline a formal evaluation framework for CALL. A formal evaluation framework based on established principles will provide more detailed and helpful information to CALL users and developers, which in turn may result in better use of CALL. The second purpose is to discover informal evaluation practices of CALL practitioners. Learning about the processes of teachers when evaluating CALL for classroom use will help direct developers of CALL materials, address user concerns, and may indirectly increase the quality of CALL materials.

Overview of the Articles

Article One. In the first article, we propose a conceptual framework for evaluation in CALL that builds on the works of Hubbard (1987, 1988, 1996, 2006, 2011), Chapelle (1999, 2001, 2007, 2010) and others (Beatty, 2010; Burston, 2003; Garrett, 2009; Levy & Stockwell, 2006; Reeder et al., 2004; Susser, 2001; Villada, 2009) and incorporates the systematic practices

from the field of formal evaluation (Davidson, 2005; Guba & Lincoln, 1981; Lincoln, 2003; Patton, 2001, 2002, 2003, 2008; Reeves & Hedberg, 2003; Scriven, 1974, 1990, 2003; Stake, 2003, 2004; Stake & Schwandt, 2006; Stufflebeam, 2003a, 2003b; Stufflebeam & Shinkfield, 2007) as well as standards from the Joint Committee on Standards for Educational Evaluation (JCSEE) (Program Evaluation Standards, 2011). The cited evaluators are prolific publishers who have influenced evaluation across domains in terms of the systematic and metaevaluative nature of modern evaluation. Throughout this dissertation, the terms *formal evaluation* and *formal evaluators* refer to the practices, standards, and ideals of the field of professional evaluation. The terms do not encompass one group or organization; however, as a collective, the organizations represented in the JCSEE are representative of the field of evaluation.

Current CALL evaluation frameworks suggested by CALL researchers and practitioners have strengths, but they often lack processes and considerations from the field of evaluation. For example, among other limitations, these frameworks neglect various stakeholders and the importance of metaevaluations. Additionally, CALL evaluation frameworks do not reference evaluation literature. CALL researchers and practitioners may be unaware of the large body of research in evaluation or they may not find it valuable or applicable to the field. Nonetheless, CALL evaluation stands to benefit from the evaluation discipline with regard to effective and systematic processes and methodological and comprehensive metaevaluation. We reference evaluation literature along with CALL literature to argue for an evaluation framework that borrows from the practices of formal evaluators and addresses the Program Evaluation Standards (2011).

We first explain the rationale for the paper and describe the methodology used in locating relevant sources. Due to the ambiguity of *evaluation*, we provide a definition for the term based

on ones similar to that of Stake and Schwandt (2006), which emphasizes evaluation as a process for determining the worth or merit of some object or activity (i.e., curricular program, software product, or learning task). We review the popular CALL frameworks and propose a framework that draws from the field of formal evaluation. We compare the framework to popular CALL frameworks and then provide a description and explanation of the processes and standards that formal evaluators tend to follow when conducting formal evaluations. These include identifying an evaluand, identifying stakeholders, determining the purpose of the evaluation, setting evaluation criteria, selecting a type of evaluation, planning and collecting data, reporting findings and implications, and evaluating the evaluation. Finally, we provide implications and suggestions for CALL in terms of increasing the quality and usability of published evaluations.

Article Two. In the second article, we examined the evaluation experiences of teachers as they selected and used CALL. For this study, we defined *expert technology users* as teachers who use technology effectively in the classroom. The research question governing the study focused on identifying the criteria teachers use in evaluating CALL. Particularly, what criteria do expert technology users implement when selecting CALL materials and activities for their classes? This entailed asking teachers the *what*, *how*, and *why* regarding the materials they use and the activities they provide for students and then drawing examples from observations of classroom practice. Knowing the factors that influence CALL selection and use can better inform developers and administrators and help them produce and support quality materials.

The theoretical perspective underpinning this multiple case research study is based in phenomenology. This emic approach lends itself to a description of particular phenomena through the eyes of the participants. Likewise, it limits the amount of external judgments made by the researcher while collecting data. Van Manen (1997) describes the researcher as the

instrument for collecting data who must engage in epoche, i.e., set aside his perceptions and judgments while collecting information. The process becomes more etic as the researcher strives to reconstruct the experiences while staying aware of his own subjectivities. The purpose of this study was to learn about how teachers select CALL materials and present their stories as cases or vignettes that exemplify the criteria they find most important.

The primary data collection method involved three interviews with each of the participants. Stake (2010) stated that interviews can help researchers find out about “‘a thing’ that [they] were unable to observe themselves” (p. 95). Each semi-structured interview lasted approximately 30 minutes, and together served as a form of triangulation. The first interview focused on the teachers’ use of technology to elicit information that enabled the interviewees to describe their experiences using technology. This interview provided a broad overview of the participants’ experiences prior to observing each teacher.

Prior to the second interview, I observed each teacher over the course of one week. Stake (2010) also suggested that exhibit questions may help the interviewees become more involved in the content of the interview. Exhibit questions can help interviewers “push respondents to sharper concentration by asking them to examine and respond to a specific statement, a story, an artifact, a quotation, or [something else]” (Stake, 2010, p. 97). Therefore, we observed the teachers three or four times during the same week, and recorded their teaching. We then reviewed syllabi and materials used by the participants and selected portions of the video recordings. Within one week of the observations we conducted a second interview with each participant. This helped direct the interview with stimulated recall and helped the teachers to focus on past activities rather than speculate about what they might have done.

The third interview was a follow-up to the previous two. Themes or topics that emerged from the previous interviews were addressed and participants could offer any other information that they felt was pertinent to the study, or correct misinterpretations of previous data. Generally the third interview took place approximately eight weeks following the observation near the end of the semester in which the teachers were observed.

After conducting the three interviews we used qualitative data analysis methods to identify themes that illustrate the criteria that teachers use to select CALL materials. We conclude by providing implications for developers, teachers, students and administrators with regard to the development, support, and use of CALL.

ARTICLE 1: A Systematic Evaluation Framework for CALL

Article Abstract

Searching prestigious Computer-assisted Language Learning (CALL) journals for references to key publications and authors in the field of evaluation yields a short list. The American Journal of Evaluation—the flagship journal of the American Evaluation Association—is only cited once in both the CALICO Journal and Language Learning and Technology (Blyth & Davis, 2007). Only two articles in these journals have cited Robert Stake, Robert Yin, Daniel Stufflebeam, Michael Scriven, or Michael Patton, five of the most influential evaluators of our generation. Prestigious journals in Computer-assisted Language Learning (CALL) lacked references to formal evaluation literature, which provides a wealth of information regarding effective evaluation processes.

We reviewed evaluation frameworks prominent in CALL and literature in formal evaluation. A comparison of the of CALL evaluation with formal evaluation showed some overlapping tasks between the two and some gaps in CALL evaluation. Practices and insights from the field of evaluation would benefit CALL researchers and practitioners with regard to conducting systematic evaluations that report evaluation findings that other researchers and practitioners find useful. The proposed evaluation framework includes common evaluation tasks conducted by evaluators in the field of formal evaluation to produce a workflow model for designing and conducting evaluations in CALL. Implications for CALL evaluators and other stakeholders indicate several areas for improvement in CALL evaluation.

Introduction

Searching prestigious CALL journals for reference to key publications and authors in the field of evaluation yields a short list. The American Journal of Evaluation—the flagship journal of the American Evaluation Association—is only cited once in both CALICO Journal and Language Learning and Technology (Blyth & Davis, 2007). Only two articles in these journals have cited Robert Stake, Robert Yin, Daniel Stufflebeam, Michael Scriven, or Michael Patton, five of the most influential evaluators of our generation. Whereas now it seems largely ignored, the field of evaluation could provide valuable insight to CALL researchers and practitioners with regard to conducting systematic evaluations that report evaluation findings that other researchers and practitioners find useful.

Chapelle (2010) stated that “the amount of published work on materials evaluation [in second language learning (SLL)] is surprisingly small in view of the impact that materials have in the instructional process” (p. 67). However, a few authors have provided guidelines for evaluation of instructional materials in SLL (Cunningsworth, 1984; Sheldon, 1988; Skierso, 1991; Tomlinson, 2003). Additionally, many prominent authors in CALL have proposed evaluation frameworks for evaluating CALL (Beatty, 2010; Burston, 2003; Chapelle, 2001, 2007, 2010; Garrett, 2009; Hubbard, 1988, 1996, 2006, 2011; Levy & Stockwell, 2006; Reeder, Heift, Roche, Tabyanian, Schlickau, & Golz, 2004; Susser, 2001; Villada, 2009). While these frameworks have their strengths, the systematic approach and practices of formal evaluators such as those of the American Evaluation Association (AEA), would help CALL evaluators design and conduct evaluations that are methodologically similar to formal evaluations, while maintaining the diversity offered by these various frameworks.

While the proposed frameworks for CALL evaluation are far from unsystematic, none explicitly describe workflows for conducting evaluations. They all implicitly mention evaluation tasks such as identifying the object of evaluation, determining the purpose of the evaluation, collecting and analyzing data, and reporting findings and implications. However, looking through the lens provided by formal evaluators, some aspects of each of these tasks are overlooked. Furthermore, the task of metaevaluation does not appear to be addressed in CALL evaluation frameworks. A framework for conducting evaluation of CALL must provide systematic steps that are based on proven evaluation practices. A systematic approach to designing and conducting quality evaluations would produce a body of transferable and usable data to help inform developers, researchers, practitioners, and students.

Prior to the Henry Ford's Model T, several inventors had developed machine-driven wheeled vehicles. Ford realized that in order to popularize such a vehicle, it needed to meet the needs of the consumers and be affordable. Soon after the success of Ford motor vehicles, others entrepreneurs began to create competing products that borrowed and improved upon Ford's technologies. Ford, in turn, made an even better product based on the consumers' interest in competing products. Today we still see this constant improvement of vehicles, yet none would argue that the Model T was ineffective or useless. Rather, the Model T was a marvel of its time and paved the way for innovation in transportation.

CALL evaluation is similar to the early pre-Model T vehicles. Evaluation in CALL has added to other fields and improved the quality of evaluation of non computer-based materials in second language learning. CALL evaluation can be further enhanced by the field of formal evaluation. In this article, we propose a conceptual framework for designing and conducting evaluations in CALL that builds on the works of Hubbard (1987, 1988, 1996, 2006), Chapelle

(1999, 2001, 2007, 2010) and others (Beatty, 2010; Burston, 2003; Levy & Stockwell, 2006; Reeder et al., 2004; Susser, 2001; Villada, 2009) and incorporates literature from the field of evaluation. The purpose of the conceptual framework is to answer the following question: What is an appropriate systematic approach for conducting evaluations in CALL?

Methodology

This review of the literature followed Machi and McEvoy's (2010) six steps for producing a literature review: selecting a topic, searching the literature, developing an argument, surveying the literature, critiquing the literature, and writing the review. When looking at literature, we decided to search for journal articles and key authors in the Linguistics/Language Behavior Abstracts (LLBA) database because it provides the richest search options for CALL when compared with Educational Resources Information Center (ERIC). Through LLBA we looked for publications containing the thesaurus entry *Computer-Assisted Language Learning* which narrowed the results to 585 articles. We further limited the search by including the term *evaluation*. From these 82 results we looked for articles that discussed evaluation in terms of proposed and actual processes. Perhaps even more helpful was looking for landmark articles that were frequently cited as well as reviewing the works of prominent CALL scholars such as Chappelle and Hubbard.

Next, we searched for literature on evaluation processes and standards. ERIC provided these three thesaurus terms: evaluation, evaluation methods, and evaluation research. A search for program evaluation standards provided several articles that address the implementation of the standards. Looking for literature regarding evaluation processes, tasks, and standards in ERIC was much less successful. Many of the articles resulting from the search were key articles in

evaluation with which we were familiar. These publications by key evaluation authors such as Patton, Scriven, Stake, and Stufflebeam proved to be the greatest source of information.

Evaluation Defined and Described

Evaluation is a systematic process that seeks to determine the value, worth, importance, significance, or merit of an object, person, or activity (Stake & Schwandt, 2006; Yarbrough, Shulha, Hopson, & Caruthers, 2010). At the core of evaluation, evaluators seek to gather and interpret data that helps us make judgments about the evaluated (Nunan, 1992). These judgments are arrived at by the interpretation of data, whether quantitative, qualitative, or both.

Systematic evaluation methodologies are similar to research methodologies. In fact, Nunan (1992) declared, “I believe that evaluations, incorporating as they do questions, data and interpretation, are a form of research” (p. 184). Levy and Stockwell (2006) referred to this as the research-evaluation nexus. However, there are notable differences in purpose between evaluation and research. The aim of most research is to produce “credible, generalizable knowledge about the nature of the world around us” whereas “evaluations help stakeholders answer specific questions or make decisions . . . [and] investigate such things as a program’s development, processes, theory, viability, outcomes and impact” (Yarbrough, Shulha, Hopson, & Caruthers, 2010, p. xxv). Research and evaluation not only differ in purpose, but they also have different audiences. Research, for example, is directed to academic peers interested in generalizable results while evaluation is directed to specific stakeholders within a specified context.

In defining evaluation, it is also important to recognize its relationship to assessment and measurement. Nunan (1992) pointed out that much of the literature confused the terms assessment and evaluation. In education, assessment is a type of evaluation that seeks to

determine the learning of students in formative or summative ways. Educational measurement is the use of assessment data to draw conclusions about students' and their abilities. In other words, assessment and measurement both fall under the broader category of evaluation. However, it is also important to note that assessment and measurement do not exclusively belong to evaluation—both are used in research as well.

The term *formal evaluation* is used throughout the article to represent the field of professional evaluation—it refers to the ideals, principles, standards and practices of *formal evaluators*. The evaluation field does not belong to one specific group or domain, rather it refers to the professional practice of formal evaluation. One example of the breadth of professional evaluation is the number of organizations represented on the Joint Committee on Standards for Educational Evaluation (JCSEE). Sponsored by more than 15 different evaluation organizations, JCSEE has published standards in program evaluation, personnel evaluation, and student evaluation. These standards are widely accepted among professional evaluators and are a driving force in the field of evaluation.

Analysis of Popular CALL Frameworks

The most popular evaluation frameworks in CALL have been posited by Hubbard (1987, 1988, 1996, 2006, 2011) and Chapelle (2001), each of which has strengths and weaknesses. In the following paragraphs we provide an overview of Hubbard's and Chapelle's frameworks. To analyze the frameworks, we took each of the principles identified in the evaluation literature as key for successful evaluations and then considered how and in what ways each principle applied to the Chapelle and Hubbard frameworks. We noted similarities and overlapping ideas. From this analysis process we created a list of evaluation principles and ideas for how they could be

realized in CALL. We then conducted a reliability check by asking two professional evaluators and two CALL scholars to review the framework for deficiencies, overlapping ideas, and clarity.

Hubbard's Framework

We will first look at Hubbard's (2011) latest publication on his framework. As with his previous publications regarding evaluation, Hubbard provides some background information about evaluation of computer-assisted language learning software in which he can situate his proposed evaluation framework.

From the onset he mentions four distinct purposes of CALL evaluation: selection for use in a course, selection for use in self-access environments or for other instructors, published reviews, and feedback during the development process. Hubbard presented these purposes not as a comprehensive list but as a specific subset to which his framework can be applied.

Hubbard narrowed the list of evaluation approaches or methodologies to three specific types: checklists, methodological frameworks, and SLA research. Checklists are essentially a combination of criteria that evaluators review and to which they assign some type of score using a Likert scale or other rating systems. While this is a common methodology used to evaluate CALL, in many cases it assumes that the evaluation criteria is *one size fits all*. Evaluators can change and alter a checklist to match the criteria specified by stakeholders, but the use of checklists as an approach or *methodology* for CALL evaluation may be confounded by its overlap with evaluation *criteria*. These checklists tend to be a list of evaluation criteria and may not provide adequate methodological concerns to CALL evaluation, omitting key procedures in the evaluation process.

The other two approaches mentioned are not exempt from similar phenomena. Methodological frameworks, as described by Hubbard, allow evaluators to form their own

questions. In this regard, the evaluator may be the sole stakeholder. This may be a limitation for evaluators with little experience who rely on CALL evaluation frameworks such as this one to guide their evaluation. Thus, it could become increasingly easy for evaluators to neglect potential stakeholders.

Hubbard (1988) based his original framework on the works of Phillips (1985) and Richards and Rogers (1982). While the confusion between criteria and methodologies in Hubbard's framework is not as prominent as it is with checklists, evaluators may interpret his suggestions as prescribed criteria. Lastly, SLA approaches seem to focus on methodologies and criteria concerning language acquisition issues. Once again, it seems as though criteria and methodologies are being grouped together. It is clear that there is a relationship between evaluation criteria and evaluation methodologies; however a versatile evaluation framework for CALL should tease apart the two in order to allow more options to evaluators when evaluating CALL. In short, evaluation criteria consists of the attributes by which the evaluand is judged and methodology refers to the approach used to learn about the evaluand with regard to those criteria.

Hubbard mentioned that his description of the framework reflected the purpose he felt most common—selection by a teacher for use in the classroom. However, he also argued that the framework could be applied to the other three purposes: selection for use in self-access environments or for other instructors, published reviews, and feedback during the development process. While possible, this may be a stretch.

Figure 1 is a diagram of Hubbard's (2011) framework. The processes Hubbard outlined in various iterations of his proposed framework include steps such as giving a technical preview, creating an operational description, considering learner and teacher fit, making appropriateness judgments, and implementing schemes.

criteria. A CALL evaluation framework should be broad enough to guide potential CALL evaluators in various situations and purposes.

Chappelle's Framework

Chappelle's (2001) framework for evaluation varies from Hubbard's. From the outset, it is clear that she did not limit the types of evaluands as strictly as Hubbard. Her framework is broad enough to consider CALL software, teacher-planned CALL activities, and learners' performance during CALL activities. She also lists standards for selecting the evaluation criteria and even suggests some specific criteria. Chappelle (2001) discussed the importance of criteria based in SLA research and stated that, "learning language potential should be the central criterion in the evaluation of CALL" (p. 52). She also lists learner fit, meaning focus, authenticity, positive impact, and practicality as criteria to be considered in CALL evaluations.

Chappelle suggests that CALL evaluations should be looked at from two perspectives: (a) a judgmental analysis of CALL software and activities and (b) an empirical analysis of the learner's performance. In many ways, this could be a recommendation for various research methodologies. She implied that the evaluand dictates, at least to some degree, the type of analysis that should be done in an evaluation. Table 1 shows three types of analyses with her suggested evaluand (object of evaluation), question, and evaluation type (method of evaluation) for each analysis.

While some CALL evaluands would appear to be best evaluated qualitatively and others quantitatively, this too may be limiting. Rather than basing the type of approach on the nature of the evaluand, it should be based on a series of factors including the nature of the evaluand, the evaluation questions, and the evaluation criteria. For example, when considering the evaluation

Table 1

Chappelle's (2001, 2007) Evaluands, Questions, and Evaluation types

Level of analysis	Object of Evaluation	Example question	Method of evaluation
1	CALL software	Does the software provide learners the opportunity for interactional modifications to negotiate meaning?	Judgmental
2	Teacher-planned CALL activities	Does the CALL activity designed by the teacher provide learners the opportunity to modify interaction for negotiation of meaning?	Judgmental
3	Learner's performance during CALL activities	Do learners actually interact and negotiate meaning while they are working in a chat room?	Empirical

of a multiple-choice test, evaluators might think that using a quantitative approach would work best.

While this may be true if the questions focus on the data from test scores and item analyses, qualitative approaches shouldn't be ignored. If the primary evaluation question focuses on learner perception of the test, qualitative methodologies may be more effective.

Hubbard (2011) and Chappelle's (2001) frameworks differ in their focus, with Hubbard emphasizing process, including its parts and the details and specific suggestions for each step. For example, he spelled out various purposes of a CALL evaluation. Chappelle on the other hand focused less on creating a procedural map for conducting evaluations and more on purpose, criteria, and methodologies.

To summarize, Chappelle's framework may be effective in certain circumstances, particularly those evaluating CALL for language issues, but it may not be helpful when considering other non-SLA issues, such as financial or hardware requirements. Following Chappelle's framework may also generate evaluations that do not consider the values of

underrepresented stakeholders such as software developers and program administrators. Evaluators who are considering non-SLA issues, such as the finances, infrastructure, or other administrative aspects would benefit from a framework that allows for other considerations such as financial issues or hardware requirements.

While the two have strengths and may be viable in certain situations, we propose a framework that is adaptable to various contexts and is dictated by the values of stakeholders, which may match values and evaluative purposes of the frameworks of Hubbard and Chapelle, but are not limited to them.

Framework for Designing and Conducting CALL Evaluations

After reviewing the popular CALL evaluation frameworks, we propose a framework that is essentially borrowed from frameworks and practices in formal evaluation. It aims to provide guidance in conducting better CALL evaluations. Its purpose is to provide direction to evaluators in conducting systematic evaluations using procedures from seasoned evaluators resulting in higher quality evaluations that are informative, efficient, useful, replicable, and to some degree transferable.

Figure 2 shows each task in relation to the others. It focuses on the careful crafting of a purpose-driven evaluation that helps evaluators identify the evaluand and stakeholders, set evaluative criteria, and determine the purpose and type of the evaluation. Performing the aforementioned tasks leads to the crafting of evaluation questions. Based on the results of previous tasks, evaluators can design the data collection and evaluation procedures, collect and analyze the data, and report the findings and implications. The rounded rectangle at the background of the figure represents *metaevaluation* and emphasizes the constant need to evaluate each task throughout the process of the evaluation. It is an evaluation task that must be

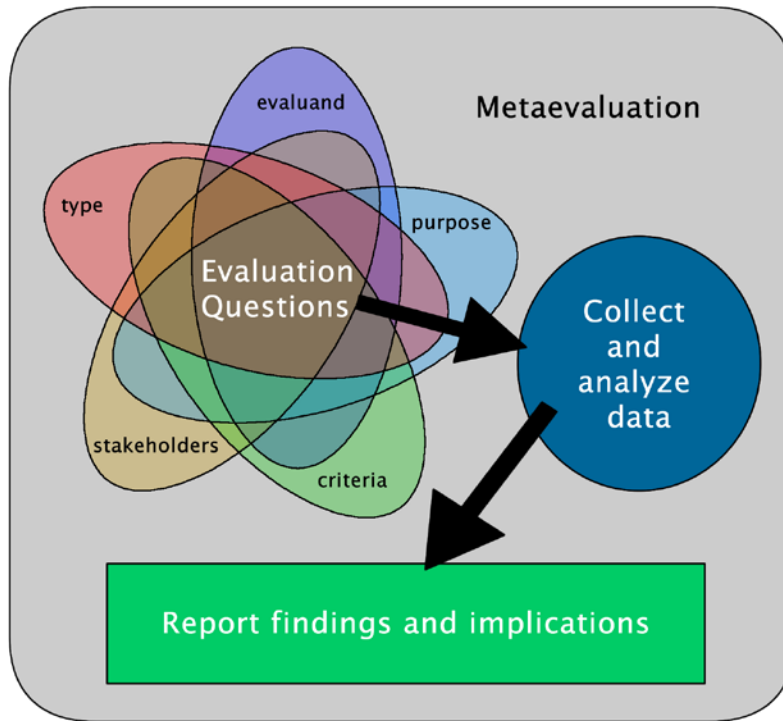


Figure 2. Framework for systematic CALL evaluation.

conducted throughout the entire process. As each task is evaluated, evaluators may need to return to previous tasks or look at and plan for future tasks as indicated.

Figure 2 also illustrates tasks for CALL evaluators. Though this is not a new framework to those in the field of evaluation, this paradigm may be new to some CALL evaluators. Rather than limit or specify the details of an evaluation, we suggest a framework that is specific enough to guide evaluators through a tried and tested process and broad enough to accommodate the evaluation of any activity or material related to CALL.

Comparison of CALL Evaluation Frameworks to Formal Evaluation Tasks

Popular CALL evaluation frameworks have many similarities and differences when compared to formal evaluation tasks. Table 2 maps formal evaluation tasks to activities mentioned in the frameworks of Hubbard (2011) and Chapelle (2007, 2011). On the surface, it seems to indicate that the nine tasks can be found in both frameworks. Essentially there are

activities that match, but there are a few differences. For example, many of the areas focused on by Hubbard were narrow and highly specific. The framework we propose is less constrained and can be applied to the evaluation of any evaluand—any CALL activity or any CALL material.

Evaluands

Hubbard claimed that his framework could be adapted toward aspects of CALL outside of courseware and websites. However, it is only focused on these two evaluands. Chapelle's suggested evaluands are placed on a continuum ranging from the evaluation of a complete course to technology used in a traditional face-to-face learning. While the spectrum allows for a wide range of evaluands, it may exclude other CALL tools or activities. The continuum focuses on technology as used in the context of a course, whether in face-to-face or online environments. Other possible evaluands may include technologies and activities that learners engage in independent of formal class settings.

Stakeholders

Although Hubbard briefly mentions developers as stakeholders when evaluating developing products, he limits stakeholders to teachers and learners. While in many cases these are the more prominent, there are always other stakeholders to consider. Each evaluation may have different stakeholders whose input should be considered. For example, school administrators, parents, and developers may be vital and perhaps underrepresented groups that can help increase the completeness as well as the validity of the evaluation. Chapelle included a lengthy list of types of evaluations and possible audiences that may also be counted as stakeholders. This list can serve as both a list of audiences to whom evaluations might be reported as well as list of possible stakeholders.

Table 2

Formal Evaluation Tasks, Hubbard's (2011) Framework, and Chapelle's (2001, 2007) Framework

Evaluation task	Hubbard's framework	Chapelle's framework
Identify the evaluand	Courseware and websites	Complete course, technology component of a course, and technology pedagogy
Identify stakeholders	Teachers and learners	Insiders (software developers, other CALL researchers), informed critics (other teachers, learners, other applied linguists), and Outsiders (other applied linguists, program decision makers, policy decision makers)
Set evaluative criteria	Technical considerations, operational description, teacher fit, learner fit	Language learning potential, meaning focus, learner fit, authenticity, positive impact, and practicality
Define a purpose	Selection for a course, selection for self-access or other instructors' use, reviews, and providing feedback for development	<i>Connected to identifying the evaluand</i>
Select an evaluation type	Grouped with data collection	Teachers' judgment, performance data, and synthesis of judgment and performance data
Develop evaluation questions	<i>Based on evaluative criteria</i>	<i>Implies the use of research questions</i>
Collect and analyze data	Checklists, methodological frameworks, SLA-based approaches	Qualitative: ethnographic and case study, interaction analysis and discourse analysis, and experimental and quasi-experimental
Report findings and implications	<i>Connected to evaluation purpose</i>	<i>Mentioned when talking about audiences</i>
Evaluate the evaluation	--	--

Criteria

One strength of Hubbard's framework is the extensive list of criteria to be considered when evaluating CALL courseware and websites. However, this is not an exhaustive list and should only be a starting point. The suggestions are specific to courseware and website evaluations, but there may be other criteria important to stakeholders. Additionally, there are no suggestions for criteria for evaluating CALL activities or even hardware used in language learning.

Purposes

While Chapelle made no explicit mention of evaluation purposes, her continuum of evaluands (ranging from the evaluation of a complete course to technology used in a traditional face-to-face learning) may indicate implied purposes similar to those Hubbard mentioned: selecting technology for a course, selection for self-access or other instructor use, and published reviews.

Evaluation Types

Unlike many of the evaluation types used by formal evaluators, Hubbard's evaluation types seemed to be tied to the data collection methods rather than the purpose of the evaluation. Hubbard (2011) cited Levy and Stockwell's (2006) three types of courseware evaluation in CALL: checklists and forms, methodological frameworks, and SLA-based approaches. Formal evaluators may adopt an evaluation type based on the purpose of the evaluation, and while some types may favor a particular type of data collection, they are not limited by it. For example, formal evaluators use quantitative, qualitative, and mixed methods to collect and analyze data gathered during an evaluation.

While Chappelle (2001) did not specify any specific type of evaluation, she did group evaluations into three distinct categories. She mentioned evaluations that are based on teachers' judgment, those that are focused on the analysis of performance data, and then evaluations that may be a synthesis of the two former types. These seem to be largely based in the participants of the evaluations—teachers and students.

Evaluation Questions

Both Hubbard and Chappelle grouped the creation of evaluation questions with other formal evaluation tasks. Hubbard suggested that the questions be connected to data collection methods while Chappelle implied the use of research questions. Both have merit and are essential areas to consider when drafting evaluation questions, but evaluation questions come through information from stakeholders and the identification of the evaluand, criteria, purpose, and type of evaluation.

Collection and Analysis of Data

As mentioned previously, Hubbard's mixed types of evaluation with types of data collection. In contrast, Chappelle (2001) suggested using research methodologies to collect data to be used in evaluations. She mentioned qualitative and quantitative approaches. She discussed the use of ethnographies, case studies, interaction analyses, discourse analyses, and both experimental and quasi-experimental methodologies. Depending on the questions asked about the evaluand, evaluators can choose a methodology that would best lend itself to the collection of valid and reliable data.

Report of Findings

In Hubbard's framework, the method for reporting findings and implications is linked to the evaluation's purpose. For example, the findings of a courseware evaluation may only be

reported to the teacher and perhaps the students. When writing a review of software or a website, the findings may be published in a journal or website. His suggested reporting venues with their strong connection to evaluation purposes were indeed helpful, but should not be seen as the only ones possible. Chapelle largely discussed her list of possible stakeholders as possible audiences, but provided little elaboration.

In addition to the suggested venues an audience for evaluation reports, there are several venues for reporting findings, and evaluators should be encouraged to find ways to share results. Because many CALL research articles focus on the efficacy of a tool or activity, they are not entirely different from an evaluation. If an evaluation uses research methods to examine the evaluand, turning a CALL evaluation into a publishable article may in many cases be possible.

The frameworks of Hubbard and Chapelle are useful frameworks that incorporate several principles as described in formal evaluation. They also provide examples or prescriptions for conducting CALL evaluation. Both may be too narrow to apply to a broad range of evaluands and evaluations. Neither emphasized the importance of a metaevaluation. The proposal to use formal evaluation principles to conduct evaluations in CALL is broad enough to capture the majority of evaluation needs of the CALL community. It also encourages delineation between various task that lead to a more focused, methodological, and systematic evaluation.

Evaluation Tasks

We've briefly mentioned the nine evaluation tasks used by formal evaluators and compared them with popular CALL frameworks. In this section we will look at each of the nine evaluation tasks. With regard to CALL evaluation, Nunan (1992) provided some guiding questions for designing evaluations that mirror the processes of formal evaluators. He suggested identifying the purpose, audience, procedures, instruments, evaluators, time frame, costs, and

reporting procedures. Formal evaluators pose similar questions that can be divided into nine primary tasks: (a) identifying the evaluand, (b) identifying stakeholders, (c) determining the purpose of the evaluation, (d) selecting an evaluation type, (e) setting evaluation criteria, (f) asking evaluation questions, (g) collecting and analyzing the data, (h) reporting findings and implications, and (i) evaluating the evaluation. In the following sections we describe each task, elaborate on the elements of each task, and provide examples from formal evaluation and CALL literature.

Before conducting an evaluation, evaluators and stakeholders work together to design the evaluation. Stufflebeam (2003) suggested that identifying the evaluand, identifying stakeholders, determining an evaluation purpose, selecting the type of evaluation, and identifying values and criteria were essential activities evaluators engage in when designing the evaluation. These are the first five tasks that should be considered when designing and conducting an evaluation.

Identifying the Evaluand

The evaluand is the object, person, or activity being evaluated. Program evaluation is often used to discuss evaluands that include educational programs, policies, or projects. In order to conduct an evaluation, evaluators need to identify what the evaluand is and what the evaluand should be. Identifying the evaluand is not always an easy process and often times there are several evaluands that can be explored. Most CALL practitioners and researchers recognize two types of evaluands: materials and activities.

Hubbard's (1988, 1996, 2006, 2011) framework focused on tools such as courseware and websites as evaluands. Levy and Stockwell (2006) suggested software, online courses, websites, computer-mediated communication, and combinations of elements (i.e., learning management

systems) as evaluands. Reeder et al. (2004) suggested three types of software that could be the object of evaluations: microcosm situations, microethnographies, and online programs.

Chapelle's (2001, 2010) guidelines for evaluating CALL tended to focus on activities as evaluands. She later identified what is taught in a complete course, what is taught through technology in a complete course, and what is taught through technology as three evaluation targets. She defined a complete course as one that is web-delivered or technology based. By looking at what is taught through technology in a complete course, she suggested that evaluators look at a subset of the total course objectives as the evaluand. What is taught through technology refers to technology that is used to enhance a traditional face-to-face classroom (Chapelle, 2007).

Because the most common frameworks for CALL evaluation focus on materials and activities, many of the published evaluations focus on these two evaluands. However, the following evaluands may also be considered when conducting an evaluation: a student using CALL, a class using CALL, a teacher using CALL, or a school using CALL. When looking at materials, evaluations do not need to be limited to software but can also include hardware. When looking at CALL activities as an evaluand, these need not be constrained by teacher-led or classroom activities but may include autonomous language learning activities. A framework for CALL evaluation should not be constrained; it should accommodate and be applicable to all types of evaluands.

Identifying Stakeholders

Stakeholders are all those who have some interest in the evaluand and by extension the evaluation. In educational settings, students, teachers, and administrators are the most frequently identified stakeholders. Often evaluators overlook curriculum and materials developers, parents, funding agencies, and other members of the community. Evaluators are sometimes overlooked

as stakeholders, but indeed become an interested party in the evaluand when conducting an evaluation (Sechrest, Babcock, & Smith, 1993). They bring expertise and unique views to the evaluation design that can help other stakeholders in the development of quality evaluations. Carefully identified stakeholders can provide essential information about the evaluand and help shape the evaluation in such a way that it becomes useful to all those invested.

Key authors in CALL have identified possible stakeholders for CALL evaluations. Levy and Stockwell (2006) discussed how designers and developers of CALL materials are evaluators. They pointed out that many of the published CALL research articles focus on tools and their evaluation. They stated, “the designer-evaluator perspective is a very important one in contemporary CALL evaluation” (p. 52). Chapelle (2001) argued that due to the complexity of CALL evaluation, all those who use CALL should be involved in the evaluation process. In a later article, Chapelle (2007) discussed audiences for evaluations. Regardless of the differences between stakeholders and audience, there is some overlap. She identified insiders, informed critics, and outsiders as three separate audiences. Insiders include software developers and other CALL researchers. Other teachers, learners, other applied linguists, and program decision makers are among the informed critics. Outsiders may overlap with critics and include program and policy decision makers. Hubbard (1988, 1996, 2006, 2011) gave high priority to teachers and learners as stakeholders.

While these prominent figures in the field have identified a myriad of possible stakeholders, many published evaluations still focus on only one or two groups of stakeholders. Villada’s (2009) proposed interpretive evaluation framework focused on multivocality as one of its main tenets. He defined multivocality as multiple voices or perspectives. Of the 24 articles on evaluations that he reviewed, 14 only addressed the perspective of the teacher. Guba and

Lincoln (1981) stated that, “the evaluator has a duty to identify all audiences, to do his best to determine what their concerns and issues are, and to honor and respond to those concerns and issues” (p. 306). CALL evaluation frameworks should incorporate the voices of all stakeholders when possible.

CALL evaluators can ensure that stakeholders are represented by asking the right questions. Guba and Lincoln (1981) suggested asking three types of questions focusing on developers, users, and nonusers of the evaluand: Who was involved in producing the evaluand, and who is using the evaluand? Who benefits from the evaluand? Who doesn't benefit or is disadvantaged by the evaluand? Asking these questions and including perspectives of the stakeholders will produce more useful and effective evaluations. These questions lead the evaluator to explore previously underrepresented groups including (but not limited to) students, nonusers, and developers.

Determining the Purpose of the Evaluation

There are various reasons for conducting an evaluation. With the evaluand in mind and in collaboration with stakeholders, evaluators should define a clear purpose for conducting the evaluation and ask detailed questions to guide the rest of the process. CALL scholars have identified some relevant evaluation purposes. Hubbard (1988, 1996, 2006, 2011) suggested the following possible purposes: (a) selection for a course; (b) selection for self-access or other instructor use; (c) reviews; and (d) feedback for development. Levy and Stockwell (2006) suggested that investigating the effectiveness of new materials as a purpose for evaluation is “one of the unique, defining features of CALL” (p. 43). They also discussed purposes such as seeing if CALL materials are working as they should, assessing value and effectiveness of CALL materials, and learning about viability and effectiveness of specific methodologies and strategies.

Their review of the literature also led them to other purposes such as assessing student attitudes and perceptions, obtaining feedback from students about CALL courses and courseware, and investigating learners' views on features of the tools they are using.

Tomlinson (2003) argued that materials evaluation “involves making judgments about the effect of the materials on the people using them” (p. 15). He continued by including an exhaustive list of possible purposes that materials evaluation seeks to measure. Here are a few that relate to CALL:

- appeal of the materials to learners,
- credibility, validity, and reliability of materials,
- ability of the materials to interest and motivate students and teachers,
- value of the materials in terms of both short and long-term learning,
- learners' and teachers' perceived value of the materials,
- flexibility of materials.

Authors of CALL evaluations have generally articulated clear purposes of their evaluations. However, their purposes have resembled those mentioned by Tomlinson (2003). Reeves and Hedberg (2003) emphasized that the purpose of evaluation is to drive decision-making. Looking at decisions that will result from the evaluation may help CALL evaluators identify purposes not previously explored that will direct the evaluation process and produce more useful evaluations.

Selecting the Type of Evaluation

The purpose of the evaluation affects the type of evaluation that should be conducted taking into consideration other variables including the skills, qualifications, or experience of the evaluators and the interests of stakeholders. Having defined purposes that were selected by

stakeholders while keeping in mind the evaluation's resulting decisions strongly influence what kind of evaluation should take place. Many CALL evaluators and formal evaluators refer to evaluations as either formative or summative. CALL evaluators have delineated other types of evaluations that are tightly connected to their evaluative purposes. Formal evaluators suggest various models for conducting evaluations. These models provide CALL with a wealth of methods and frameworks for conducting evaluations for unique purposes. As CALL evaluators shift their independent evaluation paradigms and adopt, or at the very least borrow from, models that formal evaluators use; the efficacy, efficiency, and quality of CALL evaluations may improve. First, we will look at evaluation types as generally discussed in SLL and CALL. Second, we will look at a few of the more prominent evaluation types found in formal evaluation literature.

Evaluation types in SLL and CALL. It is important to understand the types of evaluations proposed in SLL and CALL materials evaluation. Tomlinson (2003) discussed three types of evaluation. As mentioned earlier, these are connected to the purposes of the evaluation. Pre-use evaluation looks at how the materials might benefit the users. Whilst-use evaluations investigate the value of the materials while they are being used. Post-use evaluation, he argued, is the “most valuable (but least administered) type of evaluation as it can measure the actual effects of materials on users” (p. 25). Chappelle (2001) described two types of CALL evaluation: judgmental and empirical. She argued that judgmental analyses examine the “characteristics of the software and the tasks” while empirical analyses are based on “data gathered to reveal the details of CALL use and learning outcomes” (p. 54). Reeder et al. (2004) promoted a similar dichotomy in evaluation calling one type introspective and the other empirical. They argued that introspective evaluations often result in completed checklists or reviews. Introspective

evaluations often use similar criteria and provide information about the material and are based on the reviewers' perspectives. Like Chapelle (2001, 2007), they argued that empirical evaluations involve looking at students in authentic situations. CALL authors tended to mix the *type* of evaluation with the methods for *collecting data*.

Evaluation types in formal evaluation. Formal evaluation literature is abundant in proposed types of evaluation. Table 3 provides brief descriptions of a selection of formal evaluation types. The purpose of describing these is to introduce specific ways to guide CALL evaluators. However, we do not provide extensive information for conducting each type of evaluation. Lynch (1996) is possibly the only author in SLL that has situated evaluation ideals in the context of language learning. He discussed the responsive model, the illumination model, and goal-free evaluation among others. Other prominent models include Stufflebeam's (2003) CIPP model that includes context, input, process, and product evaluations. Reeves and Hedberg (2003) discussed effectiveness evaluation and impact evaluation. Patton has also introduced utilization (participant-oriented) evaluation (2003, 2008) and developmental evaluation (2010).

Examples of formal evaluation types used in CALL evaluation. Several of the proposed purposes for evaluation of CALL correspond well to the types of evaluations that formal evaluators use. Although we provide only a small description of each evaluation type, we also suggests some scenarios in which each evaluation may be appropriate in the context of CALL. We elaborate on a few of the types and situations in the following paragraphs.

Responsive evaluation. For example, Stake's (1975, 2003, 2004) responsive evaluation may be suited to the evaluation of prototypes used in teaching. The nature of changing values and concerns of stakeholders can be difficult to account for, but Stake's model may be helpful in dealing with such issues.

Table 3

Selection of Formal Evaluation Types, Descriptions, and Possible CALL Applications

Model and author	Description	Possible CALL application
Responsive Evaluation Stake (1975, 2003, 2004)	Focuses on adapting the evaluation to changing, diminishing, or emerging concerns and issues of stakeholders.	Evaluation of a CALL resource or activity in response to concerns from students, teachers, or administrators.
Illumination Evaluation Parlett and Hamilton (1974)	Discovers and describes the underlying principles and issues of the evaluand.	Evaluation of a CALL resource or activity prior to use in a class, lab, or program.
Goal-Free Evaluation Scriven (1972)	Evaluators work independently of evaluand users to determine what the evaluand actually is or does instead of determining whether it meets goals and objectives.	Evaluation of CALL tools independent of learning outcomes. (i.e. software reviews in journals)
Effectiveness Evaluation Reeves and Hedberg (2003)	Determines if the evaluand is reaching short-term goals or objectives.	Evaluation of learning outcomes from use of CALL during the course of a semester.
Impact Evaluation Reeves and Hedberg (2003)	Used to determine if what is learned through the evaluand is actually transferred to its intended context.	Evaluation of language skills acquired via CALL in comparison to actual language proficiency gain.
CIPP Model Stufflebeam (2003)	Uses the core values of stakeholders to evaluate goals (context), plans (input), actions (process), and outcomes (product) of the evaluand.	Comprehensive evaluation that considers the context, use, and outcomes of a CALL tool.
Utilization-Focused Evaluation Patton (2003, 2008)	Focuses on intended use of the evaluation by the intended users of the evaluation.	Evaluation of CALL designed and conducted for use by program administrators or other decision makers.
Developmental Evaluation Patton (2010)	Used to evaluate innovative evaluands and adapts to issues in complex environments.	Formative evaluation of CALL software or hardware during the development process.

Illuminative evaluation. What Chapelle (2001) and Reeder et al. (2004) called *judgmental* or *introspective evaluation* may be a closer resemblance to illuminative evaluation. However, this type of evaluation is not limited to materials as evaluands and can help evaluators conduct judgmental or introspective evaluations of activities because it also provides for the exploration of student tasks and experiences (Parlett & Hamilton, 1976).

Utilization-focused evaluation. Utilization-focused evaluations may have a unique fit in CALL. As academic journals increasingly publish evaluations of CALL materials, editors and publishers should consider how these evaluations will actually be used by the stakeholders. Patton (2003) stated that, “Utilization-focused evaluation is concerned with how real people in the real world apply evaluation findings and experience the evaluation process. Therefore, the focus in utilization-focused evaluation is on the intended use by intended users” (p. 223).

Developmental evaluation. The developmental evaluation model, with its affordances for complexity, is also well-suited for CALL, which in itself incorporates the complexity inherent in technology use and language learning. Perhaps one application of this model could be throughout the development process of new products and even changing and evolving language learning curricula that employ or rely heavily upon CALL.

These are only a few of the models that formal evaluation literature has to offer CALL. When CALL evaluators consider the various types of evaluation at their disposal, they can take advantage of tried practices by experienced evaluators, which will strengthen their evaluative skills and the evaluations that they produce. The type of evaluation is connected to its purposes and affects the questions, criteria, methods, and reporting of findings of the evaluation.

Setting Evaluation Criteria

Levy and Stockwell (2006) stated, “the nature of the object of the evaluation is important in choosing suitable criteria” (p. 71). With a clear understanding of the evaluand and the purpose of the evaluation, evaluators select criteria or standards by which to judge the evaluand. Criteria should reflect the values of stakeholders and be linked to the purpose of the evaluation and the questions you are asking about the evaluand. For example, administrators may consider low operating costs and ease of teacher adoption to be important. These criteria would be considered when determining the merit or worth of the evaluand. At the core of determining criteria, evaluators should consider the evaluand and its features in conjunction with the intended purpose of the evaluand and the values of stakeholders.

The two prominent frameworks by Hubbard (1987, 1988, 1996, 2006, 2011) and Chapelle (2001, 2007, 2011) provided for limited suggestions for determining criteria by which to judge their proposed evaluand. These frameworks and other evaluations failed to look at the nature of the evaluand, its intended purposes, and their relationship with the values of stakeholders. Hubbard suggested looking at technical considerations, operational descriptions, teacher fit, learner fit, and implementation. Burston’s (2003) suggestions mimicked Hubbard’s, but he also suggested that all software be both pedagogically valid and adaptable to the curriculum as well as efficient, effective, and pedagogically innovative. Both Hubbard and Burston inferred that these criteria are sufficient for evaluation of software. While they may serve as a starting point, evaluations that follow these guidelines lack the consideration of stakeholder values and may fail to address the intended outcomes of the evaluation.

Chapelle (2001) emphasized that “evaluation criteria should incorporate findings and theory-based speculation about ideal conditions for SLA” (p. 52). As mentioned earlier, the

focus of her evaluation framework was actual CALL tasks. Language learning potential, learner fit, meaning focus, authenticity, positive impact, and practicality are the six criteria she recommended. While these overlap with some of Hubbard's (1987, 1988, 1996, 2006, 2011) criteria, the focus revolves more around language learning. With regard to this, Reeder et al. (2004) argued that there is a lack of identified criteria that addresses both learning outcomes and learning processes. They also stated that evaluative criteria often fail to connect to design and instructional methodologies. While Chapelle's (2001) criteria may help evaluators and stakeholders determine their own evaluative criteria, selecting criteria based on the values of evaluators is curiously absent. Considering the desired outcomes of the evaluand and incorporating them with stakeholders' values may address the concerns of Reeder et al. (2004).

In SLL, Tomlinson (2003) who referred to materials evaluation generally and not specifically to CALL, identified 19 principles for materials development (p. 21-22). These principles are founded in research and his own experience. These work well as criteria in the materials' subsequent evaluation. One possible strength of this list is that it may be representative of the values of CALL practitioners and stakeholders alike.

Obviously, evaluating materials to ensure that they meet all possible criteria would be a large undertaking and may not even be practical, but the point is clear. Evaluation seeks to determine what should be. By clearly articulating the criteria or standards by which the evaluand will be measured, evaluators can have a directed study that lends itself to clear and defensible results, leading to clear and defensible decisions. Only by considering the intended evaluand, outcomes, and stakeholders' values can effective criteria be selected and used in the evaluation process.

Developing Evaluation Questions

Following the previous evaluation tasks, evaluators and stakeholders should ask questions. Figure 3 demonstrates the relationship this task has with the previous ones. With a purpose and type of evaluation in mind, a clearly identified evaluand, and set evaluation criteria, asking questions about the evaluand should be a well-informed task.

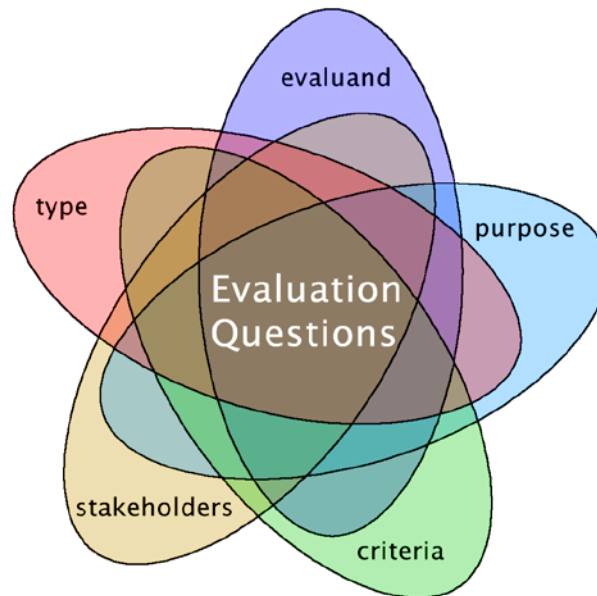


Figure 3. Evaluation tasks with the development of evaluation questions.

Stake (2010) emphasized the importance of asking questions before selecting methods to collect data. However, he also pointed out that in some situations you may define evaluation or research questions, select methods to collect data, and then return to the question in order to tweak it to work better with the chosen methods. Figure 3 shows that each of the previous tasks are intertwined and culminate in the questions asked about the evaluation—questions that drive the evaluation.

Similarly, evaluators may need to return to the previous evaluation tasks after asking the research or evaluation questions. In cooperation with stakeholders, evaluators may need to

reconsider evaluation criteria and make any necessary changes to the criteria or the question to make sure that each part of the evaluation matches the others.

In order to have questions that match the evaluation, evaluators must work with stakeholders to help arrive at overall evaluation questions. What do we (the evaluators and stakeholders) want to know about the evaluand? What questions do we need to ask to achieve our purpose? In many ways, the process of determining appropriate and evaluable questions resembles the same process that researchers undertake when asking research questions. It is essential that evaluators and stakeholders realize the connectivity among the first five evaluation tasks and developing evaluation questions. In concert, those five tasks help the development of effective evaluation questions, which in turn may require returning to those same five tasks.

For example, personnel in an intensive English program (stakeholders) use a mobile application for L2 literacy (evaluand), and want to see if it leads to an increase in reading fluency over the course of a semester (criteria) to determine if they should continue using the application (purpose). The evaluators and stakeholders decide that an effectiveness evaluation (type) is best suited for this situation. With this information they develop the main evaluation question: What effect does the mobile application have on the reading fluency of students?

Collecting and Analyzing Data

With clear evaluation questions, evaluators can design the data collection procedures and begin collecting data. Once again, the nature of the evaluand, the purpose of the evaluation, the criteria, and the evaluation questions are paramount in determining the data collection methods. The epistemological preference of the evaluator and stakeholders may inform the methodologies to be used and the types of data that are collected.

As mentioned earlier, CALL researchers and evaluators often divide evaluation into two groups: judgmental (also referred to as introspective) and empirical (Chapelle 2001, 2007; Reeder et al., 2004).

Judgmental evaluation. Evaluators commonly use checklists in judgmental evaluation of CALL materials. However, Susser (2001) suggested that the use of checklists for CALL and SLA evaluation have been criticized. He identified several areas in which various scholars described the shortcomings of checklists. Decoo (1994) questioned the accuracy, compatibility, and transferability of checklists. Squires and McDougall (1994) suggested that checklists may place too much focus on the technical features of courseware and neglect to consider pedagogical concerns. Susser (2001) referred to several experimental studies that questioned the objectivity of checklists. Similarly, he argued that some checklists may favor a particular theory of language acquisition or computer-assisted instruction. He also pointed out that the background knowledge of those completing the checklist may affect the criticality of the reviewer and the overall accuracy of the completed checklist.

Despite these arguments against checklists, Susser (2001) defended their use and explained that if they are used in the right contexts for the right purposes, checklists can be an excellent tool. Chapelle (2001, 2007), Hubbard (1988, 1996, 2011), and Reeder et al. (2004) all acknowledged value in checklists for judgmental purposes, or in other words for describing CALL materials and activities.

Empirical evaluations. Chapelle (2001, 2007) advocated that quantitative, qualitative and mixed methods approaches can be used in empirical evaluations. She connected research methodologies to evaluation. She discussed that much of the research in CALL is based on theoretical comparison research that employs experimental and quasi-experimental designs. A

theoretically motivated approach to CALL evaluation may include qualitative methodologies such as ethnographies and case studies in addition to the more quantitative experimental and quasi-experimental methodologies. She also indicated that interaction and discourse analysis may be effective methodologies used in evaluation. Levy and Stockwell (2006) also mentioned how research methods may be used in evaluation when describing the evaluation-research nexus.

Reeder et al. (2004) pointed out that many evaluations are based on experimental designs and that these might be problematic or at least make it difficult to determine the validity of outcomes from treatment groups. Reeves et al. (2005) argued that design research might be more effective than traditional experimental research. For example, experimental designs often do not control for all the variables and treatment groups are often unnatural and lack resemblance to a typical classroom. They also noted that when questionnaires are used to collect data from control and treatment groups, they are often narrow in their scope and any learning results not predicted and indicated in the questionnaire may be missed. Quantitative data collection for evaluations is not necessarily inappropriate, and in many cases may be the best data collection method depending on the evaluand, the purpose of the evaluation, the evaluation criteria, and the evaluation questions. For example, evaluators might consider student learning as criteria for and evaluation of a course or course materials. Quantitative data collected through student assessment may provide better data and more reliable indications of learner success. On the other hand, evaluators may consider student and teacher perspectives as evaluation criteria for the same course or course materials. Qualitative data collected through semi-scripted interviews may provide the evaluators with additional information. Reeder et al. (2004) suggested that open-ended questions and semi-structured interviews may provide additional insight and paint a clearer picture of the evaluand and answer evaluation questions.

Reporting Evaluation Results

Just as in research, evaluation results need to be reported. Unlike researchers who publish their work in books and academic journals, with the intention of providing generalizable or transferable findings that add to the body of research, evaluation reports might only be made available to the stakeholders. While in certain contexts a complete report with all the details may be appropriate, for many stakeholders, a brief report that includes a summary of the evaluation is often preferred. Because evaluations are extremely context dependent, they are not generalizable; their findings may, however, be transferable. Any insights gained from a specific evaluation could be used to direct evaluations of other programs and some implications might be beneficial.

Sharing evaluations with other teachers, students, administrators, and additional interested parties can be helpful as it may provide insights to others beyond the stakeholders and those who commissioned the evaluation. In language programs, evaluations could be made available to others working in the program. Software evaluations are often published in journals in both print and online formats. Additionally, evaluations of CALL tools and activities could be posted to a website making them available to a wider audience. Because evaluations are not intended to be generalizable to all contexts, published evaluations should include sufficient information such as a detailed description of the evaluand, the purpose of the evaluation, and the criteria by which the evaluand was evaluated so that others can understand the context for interpreting evaluation results and can apply the findings to their own circumstances as they feel appropriate.

Evaluations are only effective if they are used. Scriven (1990) argued that software developers and users need, “to be willing and able to do and use evaluation” (p. 3), and Patton

(2003, 2008) emphasized a utilization-focused evaluation approach for making evaluations useful to stakeholders. Beatty (2010) provided insight into the actual use of evaluations. He suggested that conducting in-house reviews by peers and making evaluations available to everyone may be one of the better approaches to creating useful evaluations. He argued that there is value in making evaluations available so they can be used by others to make conscientious decisions about the technology used in the classroom. He suggested that students might evaluate software in a self-access learning center that could be posted and made available to other students and teachers.

Many of the published evaluations in CALL journals are in the form of software reviews or research articles exploring software developed by the researcher. The target for these publications tends to be those who read the journals, and research reports and evaluation reports differ in purpose. Thus CALL frameworks of evaluation and instruction given to a prospective evaluator should stress the importance of working with stakeholders and producing effective reports to help them understand the evaluation and guide resulting decisions. Such decisions may be contextually specific and inappropriate to publish in research or other widely available reports.

Evaluating the Evaluation

Stufflebeam (1974), Scriven (1969), and others have suggested that a metaevaluation—an evaluation of the evaluation—be included in the evaluation design. Whether done internally or externally, metaevaluations not only provide a type of validation of the evaluation in question, but they also help evaluators conduct the evaluation with clear goals and systematic procedures that often lead to better evaluations. The metaevaluation should be conducted throughout the process and not just following the completion of all the evaluation tasks. To this end, the Joint

Committee on Standards for Educational Evaluation has outlined program evaluation standards (Program Evaluation Standards, 2011) that can be used to conduct metaevaluations and also help evaluators focus on the most essential aspects of evaluation. Five areas are emphasized in these standards: utility, feasibility, accuracy, propriety, and evaluation accountability.

Discussion

Implications

The proposed framework for conducting evaluations has several implications for CALL evaluators. Many of these mark a significant change from current practices. For example, the suggested framework includes explicitly setting and reporting evaluation criteria and conducting a metaevaluation. Other evaluation tasks outlined may serve to reinforce already accepted practices such as identifying stakeholders and including them in designing the evaluation. These implications may also be helpful for stakeholders other than the evaluator. Stakeholders who are well informed about conducting reliable and valid evaluations can help evaluators conduct well-designed evaluations. Administrators, publishers, designers, and others should strongly consider the following implications. Doing so will result in evaluations that are more systematic, thorough, and useful. Applying the evaluation tasks outlined in this paper might lead to the following suggestions for improving CALL evaluations. The list includes several variations from common trends in CALL evaluation to incorporate more practices from formal evaluators.

1. Evaluators should not limit themselves to CALL materials and activities as evaluands.
2. Evaluators have the responsibility to identify and determine the concerns or perspectives of all possible stakeholders. Looking at only one group (i.e., teachers, students, or developers) may not address all the issues considered in the evaluation.

3. Evaluators need to include stakeholders in articulating the purposes of evaluations. They need to consider their values and concerns.
4. Evaluators should rely on the values of stakeholders as well as research to establish criteria used to evaluate the evaluand.
5. Evaluators should use research methods to conduct dependable and systematic evaluations.
6. Evaluation reports (publications, software reviews) should be useful to their audiences.
7. Evaluators should continually evaluate their own evaluations throughout the entire process.
8. Academic journals that publish CALL evaluations should adhere to evaluation standards such as the Program Evaluation Standards (2011) or develop their own to guide evaluators throughout the evaluation process and help determine the merit of worth of other evaluations.

In addition to these implications, CALL evaluators should make an effort to incorporate formal evaluator practices in their evaluation projects. While the framework is simple and the description limited, evaluators who use this in their evaluations will produce higher quality evaluations and benefit CALL.

Suggestions for Future Research

From here, there are several questions regarding CALL evaluation that may need to be addressed. First, how might this proposed framework benefit CALL publications including peer-reviewed research and software reviews? Many publications in top tier CALL journals publish peer-reviewed research regarding the efficacy of author-generated CALL materials or CALL

activities. While some may argue that evaluation and research are similar, future research and initiatives regarding appropriateness of such publications would be helpful. Or in other words, should research that is essentially evaluation be portrayed as research? How does the evaluation of author-generated products benefit CALL in terms of the body of research it aims to contribute to?

Similarly, the current conventions for software reviews in these same top-tier journals need to be evaluated. How effective are these published reviews? As outlined in the article, our proposed framework for CALL evaluation includes essential tasks that are missing from popular CALL evaluation frameworks. Software reviews that follow our proposed framework will provide more information to readers and make them more readily usable by those same readers.

Separate from the previous questions regarding the role of evaluation in published literature, CALL evaluators should consider the used of evaluation standards. Is there a need for standards or guidelines similar to those proposed in the field of formal evaluation or would currently adopted formal evaluation standards such as the Joint Committee on Standards for Educational Evaluation be sufficient for CALL evaluations? Regardless of the answer to this question, it needs to be asked and studied, and the field of CALL should look to the field of evaluation for guidance and understanding.

Conclusion

In this paper, we have reviewed the popular CALL frameworks and formal evaluation tasks and illustrated the gap between formal evaluation and CALL evaluation. Our proposal to implement formal evaluation practices into CALL evaluation may help provide evaluations that address several issues that have been overlooked in CALL evaluation. The field of CALL needs to be more aware of the practices in mainstream evaluation and apply them when evaluating

CALL materials and activities. Formal evaluators and publications have much to offer CALL. There are a plethora of principles, ideals, and practices from which CALL evaluation may benefit. CALL evaluation publications should reflect the expertise that experienced evaluators bring to evaluation and be based on principles similar to those to which formal evaluators espouse.

Article References

- Beatty, K. (2010). *Teaching and researching: Computer-assisted language learning* (2nd ed.). London, England: Pearson Education Limited.
- Blyth, C., & Davis, J. (2007). Using formative evaluation in the development of learner-centered materials. *CALICO Journal*, 25(1), 1–21.
- Burston, J. (2003). Software selection: A primer on sources and evaluation. *CALICO Journal*, 21(1), 29–40.
- Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge, England: Cambridge University Press.
- Chapelle, C. (2007). Challenges in evaluation of innovation: Observations from technology research. *Innovation in Language Learning and Teaching*, 1(1), 30-45.
- Chapelle, C. (2010). The spread of computer-assisted language learning. *Language Teaching*, 43(1), 66-74.
- Cunningsworth, A. (1984). *Evaluating and selecting EFL teaching materials*. London, England: Heinemann.
- Decoo, W. (1994). In defense of drill and practice in CALL: A reevaluation of fundamental strategies. *Computers & Education*, 23(1-2), 151–158.
- Garrett, N. (2009). Computer-assisted language learning trends and issues revisited: Integrating innovation. *The Modern Language Journal*, 93 (focus issue), 719-740.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass Inc.

- Hubbard, P. (1987). Language teaching approaches, the evaluation of CALL software and design implications. *Modern Media in Foreign Language Education: Theory and Implementation* (pp. 227-254). Lincolnwood, IL: National Textbook Company.
- Hubbard, P. (1988). An integrated framework for CALL courseware evaluation. *CALICO Journal*, 6(2), 51–72.
- Hubbard, P. (1996). Elements of CALL methodology: development, evaluation, and implementation. In M. Pennington (Ed.), *The Power of CALL* (pp. 15-32). Houston, TX: Athelstan.
- Hubbard, P. (2006). Evaluating CALL software. In L. Ducate & N. Arnold (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching* (p. 313–318). San Marcos, TX: CALICO.
- Hubbard, P. (2011). Evaluation of Courseware and Websites. In L. Ducate & N. Arnold (Eds.), *Present and Future Promises of CALL: From Theory and Research to New Directions in Foreign Language Teaching*, Second Edition. San Marcos, TX: CALICO.
- Levy, M., & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer-assisted language learning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Lynch, B. K. (1996). *Language program evaluation: Theory and practice*. Cambridge, England: Cambridge University Press.
- Machi, D. L. A., & McEvoy, B. T. (2008). *The literature review: Six steps to success*. Thousand Oaks, CA: Corwin Press.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge, England: Cambridge University Press.

- Patton, M. Q. (2003). Utilization-focused evaluation. *International Handbook of Educational Evaluation*, 9(1), 223–244.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY: The Guilford Press.
- Reeder, K., Heift, T., Roche, J., Tabyanian, S., Schlickau, S., & Golz, P. (2004). Toward a theory of e/valuation for second language learning media. In S. Fotos & C. Browne (Eds.), *New Perspectives on CALL for Second Language Classrooms* (p. 255–278). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Reeves, T. C., & Hedberg, J. G. (2003). *Interactive learning systems evaluation*. Englewood Cliffs, NJ: Educational Technology Publications, Inc.
- Reeves, T. C., Herrington, J., & Oliver, R. (2005). Design research: A socially responsible approach to instructional technology research in higher education. *Journal of Computing in Higher Education*, 16(2), 96–115.
- Richards, J. C., & Rodgers, T. (2012). Method: Approach, design, and procedure. *TESOL Quarterly*, 16(2), 153–168.
- Sechrest, L., Babcock, J., & Smith, B. (1993). An invitation to methodological pluralism. *American Journal of Evaluation*, 14(3), 227–235.
- Scriven, M. (1969). An introduction to meta-evaluation. *Educational Products Report*, 2(5), 36–38.
- Scriven, M. (1972). Prose and cons about goal-free evaluation. *Evaluation Comment*, 3(4), 1–7.

- Scriven, M. (1990). The evaluation of hardware and software. *Studies in Educational Evaluation, 16*(1), 3–40.
- Sheldon, L. E. (1988). Evaluating ELT textbooks and materials. *ELT Journal, 42*(4), 237.
- Skierso, A. (1991). Textbook selection and evaluation. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 432–453). Boston: Heinle and Heinle.
- Squires, D., & McDougall, A. (1994). *Choosing and using educational software: a teachers' guide*. Philadelphia, Pennsylvania: Routledge.
- Stake, R. E. (1975). *Program evaluation, particularly responsive evaluation*. Kalamazoo: Western Michigan University Evaluation Center.
- Stake, R. E. (2003). Responsive evaluation. In E. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (Vol. 9, pp. 63–68). Springer International Handbooks of Education.
- Stake, R. E. (2004). *Standards-based and responsive evaluation*. Thousand Oaks, CA: Sage Publications, Inc.
- Stake, R. E., & Schwandt, T. A. (2006). On discerning quality in evaluation. *The Sage Handbook of Evaluation, 404–418*.
- Stufflebeam, D. L. (1974). Alternative approaches to educational evaluation: A self-study guide for educators. In J. Popham (Ed.), *Evaluation in education: Current applications* (pp. 97-143). Berkley, CA: McCutchan Publishing.
- Stufflebeam, D. L. (2003). The CIPP model for evaluation. *International Handbook of Educational Evaluation, 9*(1), 31-62.
- Susser, B. (2001). A defense of checklists for courseware evaluation. *ReCALL, 13*(02), 261–276.

- Tomlinson, B. (2003). Materials evaluation. In B. Tomlinson (Ed.), *Developing materials for language teaching* (pp. 15–36). Trowbridge, Wiltshire: Cromwell Press.
- Villada, E. G. (2009). CALL Evaluation for early foreign language learning: A review of the literature and a framework for evaluation. *CALICO Journal*, 26(2), 363-389.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The Program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

ARTICLE 2: Criteria Language Teachers Use When Selecting CALL Technologies

Article Abstract

Identifying the criteria that language teachers consider when selecting technologies may help inform software and hardware designers and developers as well as program administrators regarding development and adoption issues of CALL technology. It may also help other language teachers curious about their own technology uses. The focus of this case study research was to look at six language instructors considered to be experts in their use of technology in the classroom and examine the criteria they use when selecting CALL resources or activities in the classroom. Interviews, recorded classroom observations, and analysis of teaching materials resulted in three themes: consideration of pedagogy, consideration of convenience, and consideration of authenticity. These are discussed in the context of language teaching, CALL materials development, and language program administration.

Introduction

Since the early beginning of computer-assisted language learning (CALL) research, emphasis has been given to the study of technology use of autonomous language learners, alternate language learning environments, types of technology used in the classroom, and even teacher perceptions of technology use. While much has been revealed about these topics, less is known about the criteria teachers use to select the technology that they utilize to enhance language learning and teaching. Language teachers are largely responsible for the use of technology in the classroom, be it in traditional language classrooms or other language learning environments. Two questions aim at revealing information regarding this criteria. Why do teachers who are expert technology users choose to use particular technologies? How do they decide when not to use a particular technology or no technology at all?

Identifying the criteria that language teachers consider when selecting technologies may help inform software and hardware designers and developers as well as program administrators regarding development and adoption issues of technology in CALL, not to mention other language teachers curious about their own technology uses. With a better understanding of such criteria, the quality of software developed may increase as designers keep the values of end users at the forefront. If technology is a tool that enhances language learning, administrators can make decisions based on the values of those who effectively use technology. Ultimately, teachers are the gatekeepers regarding technology used in the classroom. When teachers are aware of the criteria used in selecting software, hardware, or any activity utilizing technology in the classroom, their ability to make informed decisions regarding technology use may increase.

The primary research question driving this study is: What criteria do language teachers consider when selecting and using technology to enhance language learning? By criteria, we are

referring to any factor that teachers consider in selecting technologies, including both barriers and motivators. Thus, we seek to understand the evaluation experiences of teachers with regard to technology use and non-use.

Literature Review

Literature on language teachers' use of technology represent various issues that may influence the use of technology in the language classroom. Teachers' education, attitudes, experiences, confidences, and perceptions all seem to contribute to the use of technology in the classroom.

Teacher Education

Egbert, Paulus, and Nakamichi (2002) asked how teachers learn about CALL-based activities, how the learning in their coursework impacts their current teaching, what factors influence computer use, and how teachers continue to informally learn about CALL. The answers to these questions inform us about factors that may influence the selection of CALL materials. They found that teachers generally learn about technology use on their own and that coursework is generally decontextualized. The strongest factors influencing computer use are lack of time and resources as well as other curricular or institutional restrictions. While these provide valuable information, the question still remains: what criteria do teachers use when selecting CALL materials?

Teacher Attitudes toward Technology

Similarly Kessler (2007) explored the attitudes of teachers toward technology and the type of CALL training they received. He described formal training as the instruction teachers might receive in a classroom setting and informal training as the education received through personal experiences and self-study. He surveyed 108 graduates with master's degrees in

teaching English to speakers of other languages (TESOL) and found that teachers are more influenced by informal training in CALL than they are by formal education. He argued that if teachers continue to learn in this informal setting they “may not be able to exploit the resources and learning opportunities available to them as CALL continues to evolve” (p. 184). In other words, according to Kessler’s research, CALL training needs to facilitate informal training and provide formal training to help teachers keep abreast of the latest developments in software and hardware that can be used in CALL.

Teacher Experience with Technology

Another factor influencing teachers’ technology use is their experience as teachers. Meskill, Mossop, DiAngelo, and Pasquale (2002) conducted interviews to understand the differences between those who use technology and those who do not. They interviewed experts and novices alike and reported that the five teachers with high expertise in CALL felt inclined to discuss how and why they used technology. Perhaps the most interesting finding reported in this study is that novice teachers who had received *cutting edge* training in computer use felt less comfortable in their integration of technology in the classroom than teachers with more teaching experience and less *cutting edge* training. While not explicitly stated in the publication, the researchers hint at the benefits of learning effective classroom practices from experienced teachers in order to better integrate technology into the language classroom.

In addition, Wetzel, Zambo, and Ryan (2007) observed K-8 classrooms of both experienced teachers and beginning teachers. They found that experienced teachers used technology more often in their classroom, but could not provide a reason for the phenomenon. They suggest that time teaching might affect the use of technology, but it may also have something to do with a combination of factors involving both experience and confidence, among others.

Teacher Confidence in Technology Use

Teacher confidence also influenced teachers' use of technology. Kessler and Plakans (2008) explored teacher confidence and CALL with regard to use and attitudes of digital audio and video. Their naturalistic approach involved logging classroom practices and interviewing seven ESL teachers. These teachers were selected not because of their expertise in CALL but because they all used technology in their classrooms. Through interviewing the participants, the researchers identified them as highly confident, contextually confident, and less confident. They found that highly confident teachers don't necessarily use digital audio and video regularly and often only do so because they are required. Contextually confident teachers tended to be more "reflective [and] cautious in use" (p. 278) of technology and integrated it more frequently in the classroom. They found that less confident teachers' use of digital audio and video resembled their prior use of analog media. The study provides useful information regarding technology confidence and use, but does not seek to understand how the teachers chose the materials that they used.

This was similar to Brantmeier's (2002) findings. She surveyed and interviewed 10 Ph.D. students in a semester-long seminar on CALL and found that instructors proceeded cautiously when using technology in their classrooms. Furthermore, teachers well-informed by second language reading and CALL theories developed positive views on technology use for reading instruction.

Teacher Perception of Computer Use in the Classroom

Finally, Kim (2008) investigated ESL and EFL teachers' perceptions of computer use in the classroom. She asserted that much of the CALL literature assumes that the benefits of technology use would promote a constructivist approach to learning rather than the typical

teacher-centered approach. After interviewing 10 teachers for 50 minutes each and using a grounded theory approach for data analysis, Kim concluded that teachers viewed the use of technology primarily as an instructional tool. While the author illuminated an important facet in teachers' experiences, the participants had similar backgrounds and abilities with technology that may not have painted an accurate picture of teacher experience at varying levels of expertise in technology. Additionally, the interviews probed their perceptions and did not focus on the choices and processes that these teachers undertook when using CALL. While not the purpose of the study, this information might prove helpful in understanding their perceptions of technology use.

In summary, these studies focused on teacher factors that influenced technology use in classrooms. If teacher education programs want future teachers to use technology effectively, the technology courses need to be hands-on (Egbert, Paulus, & Nakamichi, 2002; Kessler, 2007; Meskil, Mossop, DiAngelo, & Pasquale, 2002). They also need to give future teachers a sound foundation in pedagogy so that technology use enhances learning (Brantmeier, 2002; Kessler & Plankans, 2008; Kim, 2008). Educators cannot depend on formal training as a determining influence in CALL usage by prospective teachers. Additionally, educators should not depend on informal training as the definitive resource for CALL education. Although researchers have investigated factors that affect technology integration in second language classrooms, such as formal and informal training, there is a dearth of research that actually explores the process or decisions teachers undertake in the selection of CALL materials and activities. There is a hole regarding the questions that teachers ask themselves when choosing to use specific CALL-mediated instruction. Learning what standards or criteria teachers use to select CALL materials and activities may be helpful to other teachers, students, administrators, and developers.

Method

The following section outlines the research design and describes the research site, the participants, the methods used, and data analysis processes.

Design

This multiple case research (Stake, 2010) is largely based on a variant of a social constructivist epistemology, but it also borrows from phenomenology. While various ideologies exist regarding the nature of truth and whether it is absolute or solely created through an individual's reality, we maintain that people's actions are based more on their perceptions of reality and the shared perception gained through reciprocal understanding of others' perceptions than they are on the actual reality or truth of the observed. As such, by understanding the perceptions of others and constructing a shared understanding of phenomena, we can better understand the nature of the observed.

The theoretical perspective underpinning this multiple case research is based in phenomenology. This emic approach lends itself to a description of particular phenomena through the eyes of the participants. Likewise, it limits the amount of external judgments made by the researcher while collecting data. Van Manen (1997) describes the researcher as the instrument for collecting data who must set aside his own perceptions and judgments while collecting information. The process becomes more etic as the researchers strive to reconstruct the experiences while staying aware of their own subjectivities. The purpose of this study was to discover the criteria language teachers consider when selecting and using technology to enhance language learning. Observing and describing each case helped us to identify criteria these expert teachers use when selecting CALL resources and activities for use in the classroom.

Research Site

Brigham Young University has a long history of language teaching and learning. The university reports that 50 languages are regularly offered to students, with 30 more taught on special request. While 94 percent of the student body is from the United States, the remaining six percent of the student body comes from 120 different countries. Despite the largely homogenous nationality of the student body, an astounding two-thirds of BYU students speak a second language. BYU also reports that 31 percent of all students are enrolled in language classes each semester (BYU, 2012).

Additionally, BYU has a strong history of support for CALL. As early as 1973, scholars at the university were studying computer-assisted instruction, and, more specifically, TICCIT (Time-shared, Interactive, Computer-Controlled Information Television), one of the most used CALL approaches of its time (Bunderson, 1974). Other highly successful CALL-specific packages, such as CLIPS (Computerized Language Instruction and Practice Software) and ELLIS (English Language Learning and Instruction System) were also developed at BYU. In 1983, Hendricks, Bennion and Larson published an article in the then newly created *CALICO* Journal describing technology and language learning at BYU. Several BYU students and faculty have published in top-tier CALL journals and have held leadership positions in various professional development organizations dedicated to CALL. With such a storied and rich tradition of language learning and CALL specifically, BYU was an ideal context for investigating teachers' criteria for use of technology in language learning.

Participants

We used extreme case sampling (Patton, 2002) to identify those teachers who effectively use technology in the language classroom. In this study, we refer to these participants as *expert*

technology users. As reported by Kessler (2007), experienced teachers tended to be the ones that felt most comfortable talking about their technology use, so we searched for participants who were experienced language teachers. Additionally, teachers who are frequent users of technology have more experiences selecting and evaluating technology. Selecting participants who are *expert* technology users and who have experience teaching language helped us to get more information regarding evaluation criteria than *novice* teachers or teachers who rarely use technology. In other words, our definition of expert and novice technology users does not refer to the number of years of experience teaching that the participant has. Our teachers all had at least one year of experience teaching, but were selected because of their perceived use and proficiency in using technology.

We first looked at the 34 languages that would be taught during the semester the research was conducted. We then looked at the faculty that would be teaching. In many instances, there may have been one to five different instructors for a particular language. In far fewer departments there were several different instructors with varying experience. We targeted those departments due to the wide variety and the likelihood that there would be a definite outlier in terms of a teacher who was an expert user of technology for language learning. These 13 languages included Arabic, American Sign Language, Chinese, English, French, German, Hebrew, Italian, Japanese, Korean, Portuguese, Russian, and Spanish.

We approached department chairs and course coordinators in the eight departments where these languages were taught and explained the proposed research. We then asked these individuals to identify the instructor for each language that was taught in their department that would qualify as an expert technology user. Most responses were instantaneous, as each department appeared to know exactly who the “technology expert” was. A few departments

chose not to participate. There was some difficulty in selecting an Arabic teacher because the department felt that their teachers were generally equal. Among Chinese teachers, the department selected three individuals. After talking with each instructor, two of the three indicated that the third would be the best example of an extreme case. She agreed to be a research participant.

Six participants were ultimately selected for this study, representing teachers of Spanish, ESL, Arabic, Mandarin Chinese, German, and Russian.

Subjectivity Statement

All members of the research team are affiliated with BYU, as are the six participants. We wanted to learn more about the perspectives of our peers when dealing with CALL and felt that the better we understood the perspectives of other teachers, the better we can develop and implement CALL materials. One member of our team had previously taught ESL with two of the participants prior to the study. Another team member taught these same two participants in a graduate course for ESL teachers.

Knowing two of the participants, and sharing affiliation to BYU, may have allowed us to have more open, candid, and comfortable discussions. It allowed us to understand their perspectives in a more formal and methodological way with data that could be analyzed and synthesized to help us all gain a greater understanding. However, while there was the possibility for open communication and the collection of rich data, some participants may have felt uncomfortable sharing their teaching experiences, which may have limited some of the data collected. Our relationship with the participants had no negative affect on their job status. All the data was kept confidential.

Procedures

The primary data collection was done through three interviews with each of the participants. Stake (2010) stated that interviews can help researchers find out about “‘a thing’ that [they] were unable to observe themselves” (p. 95). Each semi-structured interview lasted approximately 30 minutes, and together served as a form of triangulation. The first interview focused on the teachers’ use of technology and elicited information that enabled the interviewees to describe their experiences using technology. These interviews provided a broad overview of the participants’ experiences and helped them articulate their teaching philosophy with regards to the use of technology and language learning. We conducted these interviews the week before observing each teacher.

The second interview was conducted after the participants were observed teaching. Stake (2010) also suggested that exhibit questions may help the interviewees become more involved in the content of the interview. Exhibit questions can help interviewers “push respondents to sharper concentration by asking them to examine and respond to a specific statement, a story, an artifact, a quotation, or [something else]” (p. 97). Therefore, prior to the second interview, we observed and recorded the teachers in one of their classes for one week and reviewed the materials used in those lessons. During the week following the observations, the participants and a member of the research team reviewed selected portions of the video recordings. These recordings and materials were used as stimulated recall to help the teachers focus on past activities rather than speculate about what they may have done.

The third interview was a follow-up to the previous two. Themes or topics that emerged from the previous interviews were addressed and participants had the opportunity to offer any other information that they felt was pertinent to the study. This interview served as member

checking (Stake, 2010) to verify the synthesis of the collected data. Each participant was sent a copy of their individual case and asked to review it prior to the interview. A member of the research team asked for clarifications or changes that would make the synthesis more accurate. We conducted these interviews about eight weeks after the second interview and toward the end of the semester during which the observations took place.

Data Analysis

Stake (2010) suggested that at the core of the data analysis researchers are analyzing and synthesizing the data, or in other words they are taking it apart and putting it back together again. He emphasizes that researchers are essentially interpreting the data. Coding consists of sorting and classifying data, which in many regards is both an analysis and a synthesis. Moustakas (1994) and others (Reid, Flowers, & Larkin, 2005; Smith, Flowers, & Larkin, 2009) provide differing yet similar suggestions for phenomenological data analysis. These include an emic description of the participants' experience followed by an interpretive step involving the etic perspective of the researcher.

While our approach to data analysis may not be stated as simply as Stake (2010) describes, it does follow processes similar to those proposed by the aforementioned phenomenologists (Moustakas, 1994; Reid, et al., 2005; Smith, Flowers, & Larkin, 2009). We broke down the data using the participants' words (emic) and reconstructed it into meaningful themes (etic). The emic approach is consistent with phenomenology as it seeks to understand the perspectives of the participants. The reconstruction into themes allowed us to *reveal* and *describe* the variation—an etic component. This combination of emic (textural) and etic (structural) approaches is then combined to provide a holistic interpretation of the key aspects of a lived phenomenon, which we present as cases or vignettes.

After each interview was completed, we coded the data using the exact words of the participants to identify key points. After the data was saturated with emic codes, adding our etic codes allowed us to group these coded terms into categories. We then put the data back together by correlating the emic and etic codes to help inform the final step: the identification of themes.

In order to ensure credibility of the analysis, we performed member checking. Hatch (2002) suggested providing the participants with a draft of the summary. The participants reviewed the summary and we discussed any discrepancies or incongruities in the new, shared perspective of the criteria used by teachers to select CALL technologies.

Vignettes

The following vignettes are used to illustrate the experiences of the participants. At the beginning of each case, we provide a brief background of the participant. While these teachers have much in common, they are all at different points in their careers as language teachers. Nonetheless, they have all been identified as expert technology users in their corresponding department. All the teachers, except for Justin, taught in classroom with an LCD projector, a computer station, and wall-mounted speakers. Justin was the only teacher with an overhead transparency machine; all other materials had to be checked out and brought to his classroom.

We have incorporated direct quotes that we have selected based on themes from the data analysis. Pseudonyms are used to preserve anonymity.

Justin

Justin is a full-time instructor at the English Language Center. Four class sessions were observed and recorded during one week of an intermediate-mid listening and speaking class. Students in the class came from various countries and spoke several different languages.

Justin made several remarks indicating that effective pedagogy was of utmost importance when either choosing whether or not to use technology in the classroom or choosing between

two technologies. Justin's main concern was for the learning that should occur. In his own words, he described his philosophy toward the use of technology in the classroom: "The software doesn't guide the learning exercise—I guide the learning exercise."

Justin preferred to use new/advanced technologies in the classroom, such as LCD projectors, but the lack of hi-tech hardware in the classroom hindered his ability to use technology as he wished, so he used overhead projectors. Justin mentioned his reasons for using an overhead projector in class. "It's much easier for me to bring transparencies than it is to bring up a projector." Justin expressed similar sentiments with regard to using the class set of iPod touch devices. Reserving, checking out, and transporting equipment took time and effort that, in many instances, was additional work that didn't result in additional benefit. He also mentioned that he used his cell phone. "I use the timer on my cellphone . . . and I don't have to bring up another device just to keep time." The convenience of using a device that he was taking to class everyday versus bringing a separate one was an important factor for Justin.

Justin took his classes to the computer lab fairly regularly. With regard to criteria he considered before selecting hardware or software for his students to use in the computer lab, he said that both his own familiarity with the technology and his students' familiarity played a vital role in his selection process. Justin commented that when introducing something unfamiliar to the students, it is important to consider the time investment in teaching them how to use the hardware or software and the amount of learning that will occur as a result of using technology. For example, he preferred the simplicity of Quicktime for audio recording. The software is limited to recording audio, video, and screencasts. He argued that the time it takes to prepare students to use it was minimal and the chances of them doing something wrong was quite low.

One activity that students engaged in at the ELC was elicited imitation (Cook, McGhee, Lonsdale, 2011). Students listened to a sentence and then repeated and recorded themselves saying the sentence. The software was then able to provide feedback and projected proficiency scores based on their repetition of several sentences. Justin said he used this software because, “it was created for that specific task.” He explained that when software is geared to a specific task it is easier for students to stay focused and it is easier for the students to use the software. There is little distraction in the software and little time needed to train students to use it.

Overall, Justin made use of various technologies both low-tech and high-tech. He was comfortable with technology usage and his ability to use it for his pedagogical purposes.

John

John is a graduate student in the Spanish and Portuguese department. We observed five class periods during the course of one week of a third year Spanish course. His students are native English speakers. During this semester he offered to participate in a pilot program that integrated an authentic television program into the curriculum. At the time we observed him, John was also working on a graduate thesis exploring social media in the language classroom.

John felt that as a non-native Spanish speaker, pulling in audio and video generated by native speakers in authentic situations was essential in enhancing his students’ learning. He also used authentic video and audio to give students insights into the culture of the target language. John summed up his view of the importance of good pedagogy and technology: “[Technology provides] really good access to authentic material. Authentic materials, from a language perspective, are what we need to get to our students.” He added that, “successful technology use depends on good pedagogy.”

In contrast to Justin's situation, John had a computer and LCD projector in the room where he taught. He used the LCD projector to play audio and video, show pictures, and do interactive activities. When we asked him why he chose an LCD projector over an overhead projector he stated that he "doesn't even know how to make transparencies." Using the computer and LCD projector was easier and more convenient for him.

During the class periods we observed, John used several videos. He reported that when he chose which videos to use, it inevitably came down to whether the video provided the language needed to meet the intended learning outcomes. He also noted on more than one occasion that all activities do not need to be based in technology, and that varying the types of activities in addition to having the learning outcomes in mind results in better teaching.

Although John made good use of the whiteboard in his classroom, we asked him why he used PowerPoint presentations to display grammar rules and examples instead of a handout, an overhead projector, or the whiteboard. John felt that the PowerPoint presentations were more helpful. Even if the grammar materials were written before hand, writing on the board would take an extra step. Additionally, John valued the time he had to prepare the presentations because it allowed him to prepare effectively and focus on teaching in class rather than simply presenting material. John found it much more convenient and helpful to students to show PowerPoint presentations in class and then post these presentations to a learning management system for the students to access it outside of class via the internet.

Through the observations and interviews, it appeared that John really only used one technology that was specifically purposed for learning language. A group of teachers, researchers, and developers were piloting the use of a website in some of the Spanish courses. The website provided authentic video from a popular Hispanic television program. Other

learning materials and assessments were available for the teachers to use. John's participation in the pilot study matched his philosophy about using authentic language in the classroom.

Moreover, he emphasized that authenticity is not just language spoken by native speakers, but language spoken to native speakers.

Shada

Shada is a native Egyptian Arabic teacher who works in the Asian and Near Eastern Languages department as an adjunct Arabic teacher. We observed a second year Arabic class that met five days each week. Two of those class sessions each week were presented in a different format involving a larger amount of students from various sections taught by other instructors. By her request, we limited our observations to the three traditional class sessions Shada was teaching.

During our observations, Shada did not use presentation software such as PowerPoint. In fact, most of the written language presented to students was written on the whiteboard. She felt it was worth taking the time to write on the board so that students can "see the handwriting." As a teacher she felt that it was important that her students become familiar with Arabic handwriting. She wanted them to be able to see how to write the characters. In this situation, Shada opted to forgo the use of technology in favor of a perceived learning benefit to the students.

The Arabic language curriculum in the department was supported in part from a textbook series. The textbooks came with a DVD that has video and audio of native speakers interacting. While Shada did not use all the available language samples, she selected videos based on their authenticity and relevance to the cultural or linguistic feature being taught in class. Rather than show the videos because they came with the textbook or because they are interesting, Shada used media only when it fit with what she was teaching. Shada added that, "sometimes showing a

video can be a waste of time if it is not serving its purpose.” When selecting recorded language samples for use in the classroom, her priority seemed to be on the learning task. While she realized that the audio samples that come with the textbooks were not necessarily authentic, Shada recognized that the scenarios and vocabulary were authentic to the tasks her students would do in class.

Hsiu Ting

Hsiu Ting is full-time faculty in the Asian and Near Eastern Languages department where she teaches several classes in Mandarin Chinese. During the semester we conducted the study, she was teaching a current events media-based class.

Two quotes from her interviews exemplify her approach to language learning and technology. “I love technology.” “[Technology] is an assistive tool. It’s not the primary purpose [of the lesson].” During the interviews, Hsiu Ting alluded to language learning principles that guided her teaching. She emphasized that pedagogy is a strong factor in predicting the outcome of technology-based activities.

When teaching a class that focused on current events, Hsiu Ting strongly believed in the use of authentic materials. While observing her class, she used several videos of authentic news broadcasts. In most cases, she would provide the class with a scaffolding exercise to prime them for the listening passage provided by the broadcast. She would play the authentic passage a few times and stop to check for comprehension both during playback and between repetitions. In some cases, comprehension questions were included on a handout or discussed prior to listening or during it.

Much like Shada, Hsiu Ting used the whiteboard to help her students learn how to write Chinese characters. “Students need to learn how to write the strokes—the particular stroke order. Sometimes they think they write it correctly, but they don’t.” Using the board to both

show students how to write Chinese characters and allow students to practice during class was more important to her than using technology that may or may not have helped students reach the same intended learning outcome.

Unlike Shada, Hsui Ting still made extensive use of PowerPoint presentations. She recognized that some students may have had trouble understanding the handwriting of others and wanted to provide those students with comprehensible input. Like John, she saw PowerPoint as a tool that aided in preparing pedagogical lessons because it helped her organize her ideas and focus on the specific language tasks the students would engage in. Hsiu Ting found it more convenient to have her materials in a digital format. She explained that it was easier to carry around a USB drive with materials for the different classes she taught than it was to lug around hard copies of all the materials.

We asked Hsiu Ting what technologies she regularly used in her teaching and one of her responses was that she used to have a class website. Upon further questioning, she indicated that dealing with website security issues and maintenance was not enjoyable so she began to use Blackboard and dropbox as ways to communicate with students and provide them with materials they needed in class. She found these alternatives to be more accessible to her students and easier for her to maintain.

Bettina

Bettina is adjunct faculty in the Department of German Studies and Slavic Languages where she teaches German classes. The semester we conducted the study, she was teaching two sections of a third year German course that met three times each week.

When asked about her approach or philosophy toward the role of technology in language classroom she said, “the more [technology] the merrier.” Bettina said that she uses technology

until “it gets in the way” of language learning. She emphasized the importance of fostering communication in the classroom and practicing the target language. She noted that sometimes “you have to turn it off” and start talking. “In a language class, you need to talk as much as you possibly can.”

Throughout the observations, Bettina made extensive use of the LCD projector to show documents, images, and videos. She felt that being able to point at a document and indicate which part of the text they were discussing helped eliminate confusion and made the class flow better. Students could spend more time in the text instead of asking where they were in the text.

In several instances, Bettina used a Microsoft Word document. She mentioned several factors that governed her decision. When using a word document instead of the chalkboard, she had her face to the class and she felt that she was in a better position to interact with the students. She also mentioned that what she wrote in a word document could be saved and shared later with students. The students might spend less time jotting down hurried notes and more time interacting with their peers and the teacher knowing that the notes on the screen would be available later.

The class that Bettina was teaching during the study had a literature component. Bettina mentioned that many of these texts were available online and that she provided students with links to the text rather than printing copies for each student. When selecting texts to be used in class, she reported that she first chose the literature and then looked to see if the text was available online. She first considered the content needed to teach students and then looked for a technology solution.

Jessica

Jessica is full-time faculty in the Department for German Studies and Slavic languages. She teaches a variety of Russian classes. We observed a first year Russian course that met daily. A teaching assistant taught two times each week. We observed the three days that Jessica taught the class.

Regarding her approach to technology usage in the language classroom, Jessica stated that “technology should always be used in pursuit of a goal and not for the sake of technology itself.” For each activity that Jessica used in the classroom, she articulated its pedagogical foundation. She could clearly explain without hesitation the pedagogical benefits of each technology she used. For example, she quickly pointed out the visual, aural, and tactile benefits when using individual whiteboards in class. Students listened to the target language and produced the target language through writing, which provided different experiences with the language that may have catered to variations in students’ learning styles.

Like Shada and Hsiu Ting, Jessica felt that providing the students with opportunities to write the characters of the language was important. She mentioned that having the students use a computer would be less helpful because the keyboard in the classroom did not support Russian language input. In this situation, only one student would be able to type at a time. Instead of using a hi-tech alternative, Jessica opted to use the whiteboard for practical and pedagogical reasons. She took this a step further in her class by bringing a small whiteboard, markers, and erasers for each student to use simultaneously to share their practice with the rest of the class. “I like that for pedagogical purposes—it gets me immediate feedback. Everybody is involved and it is multi-sensory.”

At the beginning of each class, she used the projector to display English words on the whiteboard next to which the students would then write the Russian translation. Because she worked with a teaching assistant for this class, the two would often need to coordinate the words to be displayed. Once Jessica knew which words to use that day, she created a PowerPoint presentation to project the words on the whiteboard during class. By the third observation, she stopped doing this activity in this fashion because the amount of effort to prepare for this activity was more than the resultant learning. Instead she shared a word document with the teaching assistant that contained an updated list of words. As class was starting, she referred to the document and wrote the words on the board. Both Jessica and the teaching assistant were able to accomplish the intended learning outcome with less effort than they had given previously.

Findings

Following the analysis method previously described we coded the data collected through interviews and observations. From this analysis we identified three emergent themes: consideration of pedagogy, consideration of convenience or accessibility, and consideration of purpose and authenticity. These themes were consistent among all participants and across all interviews.

Consideration for Pedagogy

The pedagogy involved in implementing technology use to learn language was a predominant theme expressed in the interviews with the participants. Justin put more focus on his students' learning and considered the convenience or inconvenience of technology use as it relates to good pedagogy. When Justin was asked to describe an experience when technology was not used well, he focused more on how "the learning objectives were not being met," rather than on the technology not working. Justin also considered if the technology would distract rather than enhance language learning. "If I think someone is going to be more interested in

watching me use the technology than actually watching or interacting with the technology in a way I'm hoping they will, then I will choose not to use it." He continues, "I try to use [technology] if I feel it will enhance their learning and provide me with an easier opportunity to give individualized feedback." As mentioned earlier, Justin considered the teacher to be essential in the effective implementation of technology. He said, "The software doesn't guide the learning exercise—I guide the learning exercise." His implication was that good teaching is defined by the teachers' actions and not the technology itself.

While John enjoyed using authentic video, he commented that when considering an authentic video or audio sample for use in the classroom, he also considers the linguistic or cultural input the media provides to students. If the media does not correspond with intended learning outcomes, he would not use it. On one occasion he had planned to use technology but did not end up doing so. He said that, "there was a disconnect in what I had prepared and what the students were understanding." In other words, the in-the-moment learning needs of these students were more important than carrying on with the planned activity that used technology.

Hsiu Ting indicated that she used technology to assist in language learning because it is effective. As we continued to talk about this, she revealed that she regularly does pre- and posttests to help gauge her students increase in language proficiency. She also took student evaluations seriously to better understand how they feel the class is contributing to their language learning. She indicated that the data speaks for itself and shows that students are learning Chinese through the use of media. She also mentioned that, "over 85 percent of the students think that this media class really builds a solid foundation for their listening and reading [proficiency]." Bettina considered teacher-student interaction and use of the target language to be extremely important. If using technology in the classroom inhibits the use of the target language

in the classroom, there may be an alternative non-technology activity that would be better suited for the language-learning task.

Before considering the use of technology in the classroom, Jessica “[thought] of the outcomes for the day, the objectives, the material that is scheduled to be covered and then consider[ed] the activities [that could be done] in pursuit of those outcomes.” She also commented on how technology can distract from the learning process. For each activity from the classes we observed and reviewed with her, Jessica was keen to provide a pedagogically sound rationale for that activity.

Although Shada did not make extensive use of technology during our observations, it was clear that her choice to use or not use technology was based on the students needs. She used technology when it would enhance learning, and opted for other materials when technology might not be helpful to students.

Consideration of Convenience or Accessibility

One criterion that the participants considered was the convenience or the accessibility of the technology. Convenience, as used here, refers to a few different concepts. First, it is used in the traditional sense—something that can be done with little effort. The inconvenience of bringing hardware to a room whose technology was limited to a chalkboard and overhead projector was a common topic mentioned in Justin’s interview. Similar issues were made apparent in the interviews with the other participants. Bettina’s decision to modify her use of the whiteboard and overhead at the beginning of class is one example. Another is John’s choice to use an LCD projector rather than an overhead projector. If something low-tech or no-tech is readily available, and can do the same job for the same or less effort, the teachers would use what was available. Many of the reasons why John used one technology over another are tied to

how convenient it was to use the technology. Hsiu Ting also mentioned that using PowerPoint was convenient and easily accessible. She could reuse and recycle presentations. She could include links to text and audio samples in the presentation, which saved her time in class.

Referring to criteria she used when choosing one technology over another, Bettina said that, “convenience is key.” Provided that her pedagogical priorities were met, Bettina considered the convenience of using the technology. Jessica explained that it is important to her that the amount of time and effort she put towards using technology resulted in increased learning. If preparing to use a particular technology took a significant amount of time or was not easily accessible and only yielded small increases in learning, she may choose not to use that technology.

Familiarity. Familiarity is a subset included in this theme of convenience and accessibility. An unfamiliar technology or technology based task was considered to be inconvenient or inaccessible. This involves both the teacher’s and the students’ familiarity with a technology used for learning. Justin, along with other participants, felt that the more familiar students and teachers were with a technology, the more convenient it was to use it over something with which they were less familiar. Justin stated, “I use [technology I am familiar with] just because I know I will be able to troubleshoot the technology as it’s being used and I know that I will be able to present the technology in a way that is the least disruptive.”

Reliability. Another subset focused on the reliability of the technology. During the interviews, each participant mentioned that technology in the classroom was of no value if it failed to function properly. Reliability was a criterion that Jessica felt was important when using technology. She mentioned that she often preferred to make online videos available offline

because the hosting website might go down or there might be internet access problems during class.

Consideration of Purpose and Authenticity

Purpose and authenticity made up the third theme. The two were grouped together because the participants used them together. The participants frequently elaborated on the purpose of a particular CALL activity or the authenticity of an activity or material. Generally speaking, their purposes of technology use focused on how well the CALL material or activity met learning outcomes by providing authentic language in semi-authentic situations.

Of all the participants, Justin made the most use of software that was designed with language learning in mind. For part of the class time during the week, he would take his students to a computer lab where students used in-house developed software and proprietary software. He found that using software that was designed for a specific language-learning task was more beneficial to his students' language proficiency development.

During the five days that we observed John, he made use of video that was available on the internet and throughout the interviews he focused on the importance of authenticity. Most notably he mentioned his preference for materials that were intended for native speakers rather than videos targeted at language learners. Hsiu Ting also felt that it was important to use authentic videos because it exposed students to native speech at various rates of speed and new vocabulary among other linguistic features they may not have been exposed to in a traditional classroom. While there is a clear contrast in Justin's reason for using software that was geared to language learning and John and Hsiu Ting's preference to materials that are not targeted at language learners, these two approaches were quite similar. They all considered the purpose of

the material being used and chose the one that they felt was most effective in the language learning process.

Shada's criteria for selecting audio and video language samples in the classroom was based on the learning needs of her students and the authenticity of the language sample. She understood the need for authentic video, but also considered video intended for language learners. While she recognized that scripted material lacked authenticity, she added that this was just one aspect of the video's authenticity. The vocabulary and situations were based on authentic language that students will encounter.

Discussion

Although the study is not without limitations, the findings have several implications on CALL and SLL.

Limitations

As with many qualitative studies, the most prominent limitation is the restrictive pool of participants. While participants were teachers of differing languages with different writing systems (among other differences), we only looked at six instructors. All of these instructors taught at the same university. While the findings of the study might be transferrable to other institutions, there is still a need to do similar studies to explore language teachers' criteria for selecting and using technology.

Another limitation related to the participants is that each one was identified as a teacher who used technology more than the other teachers in their departments. While these teachers all had similar criteria they considered when using technology it may not be representative of teachers who regularly or rarely use technology.

Implications

As mentioned in the introduction, the implications from this study affect several different groups. This section will focus on implications for administrators, teacher educators, CALL software developers, and language teachers.

Implications for Administrators. Egbert, Paulus, and Nakamichi (2002) reported that the lack of time and resources effected teachers' use of technology. At first glance the relationship between their study and ours may not be readily apparent. The teachers in this study valued the convenience in their selection of materials for use in the classroom. Lack of time to effectively use technology is an issue of convenience as is the lack of available resources.

Make technology accessible. Perhaps one of the most profound implications from the study is that administrators make technology available to teachers. The clearest example of this is Justin's inconvenient access to technology. He indicated that he would use an LCD projector more often if it were readily available in the classroom where he taught. As it is, the effort to bring the technology to the classroom and the disruption setting up the technology was not worth the effort. Looking at the other five participants, each had a projector and computer station in the classroom where we observed them teach. Each of them used the computer and LCD projector in unique ways that were not just slideshows with bullet points and text.

The example of the accessibility of an LCD projector is most likely an indicator that expert technology users would use other technologies if they were readily accessible. Smartboards, document cameras, and clickers are just a few technologies that these teachers might use. At the time of this study, tablet computing devices are penetrating the market like never before and may become more widely used than laptop computers. Administrators should

carefully evaluate the effectiveness of such devices for language learning and make provisions for teachers to have easy access to them.

Program administrators, as well as teachers, should be aware of another technology that in many classrooms is readily accessible—smartphones. If accessibility is truly a concern, then teachers and administrators should make use of readily available technology. Justin used his phone as a timer. Having speakers to connect phones to for audio playback or other connections options such as mini HDMI, AirPlay, or other wireless video connectivity could be a strength in the use of mobile technology for language learning.

Provide training. Per our findings, the above recommendation comes with the following absolutely essential caveat: administrators should make sure teachers have time available to learn and use technology. In every case, pedagogy and student learning needs trumped convenience and access. Simply putting more technology in the classroom will result in higher use, unless pedagogy is first considered. All six cases expressed the importance of convenience with regard to using technology. Familiarity plays an important role. Administrators should provide training for teachers on the use of technology and make sure they have ready access to that technology to learn and play with the software and hardware. This will allow teachers to become familiar with the technology and promote its use in the classroom.

Incidentally, the students' familiarity of the technology was also a concern for some of the participants. Providing access and encouraging use of the technology outside of the classroom may prove to increase the effective use of that technology in class.

Make informed decisions. Additionally, administrators should consider pedagogy, convenience, and authenticity when purchasing software for their institutions or mandating that a particular technology be used. If administrators of a language program decide to purchase

hardware or software, they should ask themselves and the teachers of the program how the purchased product will meet the learning needs of the students. How accessible and convenient is it for teachers and students to make use of the proposed product? For example, if teachers feel that the technology does little to help increase the language proficiency of the students, they may choose not to use it. If a language learning application is purchased that will only work on a limited number of devices and makes it difficult for teachers to have access to it, teachers may not use the application as often as intended. In short, administrators should explore what their teachers, students, and curriculum need with regard to the purpose of the software or hardware.

Teach principles of good pedagogy. Based on our observations, these expert technology users tended to be good teachers. Administrators should consider teacher training and teacher proficiency of utmost importance. If program administrators want their teachers to use technology effectively, they should ensure that teachers are aware of the pedagogical foundations of language teaching. Administrators cannot expect that the use of technology will lead to better teaching. They should remember that pedagogically sound teaching can lead to effective use of technology.

Teacher Educators. Perhaps the implication for teacher educators is more apparent than those of program administrators. Those who are involved in training new teachers should take note of these three themes. In the studies cited previously, the experiences in the teacher education program were critical in the use, misuse, or nonuse of technology in the classroom (DiAngelo & Pasquale, 2002; Egbert, Paulus, & Nakamichi, 2002; Kessler, 2007).

Provide experiences with technology. The theme of convenience has a sub-theme of familiarity. If teacher educators feel that it is important to use technology in the language classroom, they need to help teachers become more familiar with the software or hardware the

students will encounter. In the example of the in-house software that Justin used, administrators would also have the role of training teachers to use this software. These trainers need to provide seasoned teachers as well as prospective teachers with the opportunity to use the technology, become familiar with it, and ultimately be able to instruct their language-learning students on how to use the technology in question.

Teach principles of good teaching. Teacher educators need to have a strong understanding of language learning pedagogy and be able to help the students under their tutelage to have the same understanding. Egbert, Paulus, and Nakamichi (2002) reported that coursework was often decontextualized but also noted that curricular restrictions inhibited technology use. Teacher educators should help prospective teachers understand how to use technology in conjunction with pedagogical principles that match imposed curricular guidelines. This was also found in the studies by Kessler and Plankans (2008) and Kim (2008), and strongly reinforced here. A connection needs to be made between the pedagogical purposes and outcomes in a language course and the ability or inability for a particular technology to aid in reaching these purposes or outcomes. Effective users of technology are pedagogically conscientious.

CALL Software Developers. On the surface it may seem that there is little in which software developers can contribute to the pedagogical use or the convenience of CALL software in the classroom. However, there are some key software features that would serve more to promote pedagogically sound language instruction and eliminate possible inconveniences.

Consider pedagogy during development. CALL developers should allow pedagogical needs to drive software development. Teachers consider how well the technology will enhance learning and help meet pre-defined learning outcomes. Developers should consult with language

teachers, researchers, and pedagogues to identify common language learning objectives that can be enhanced through technology. Doing so during pre-development and design phases of the software will produce a tool that teachers are more likely to use. Granted, when CALL software is being developed in-house or by teachers for their own students, the software can be more specific with a narrow purpose in mind. However, CALL developers who create software for various organizations need to consider the need of their possible clients and in some cases need to ensure that their software is general enough to meet basic or common needs of the majority of language learning programs.

Ensure that software is accessible and useable. Accessibility and usability are two very important features that CALL developers should consider. This also fits well with report given by Egbert, Paulus, and Nakamichi (2002) that lack of time was an inhibiting factor in CALL implementation. Developed software should require little time to learn to use and little time to teach students how to use. It should be convenient to use with regard to the time commitment required. In today's world, we see a plethora of electronic devices, the smallest of which are more powerful than those available ten years ago. In universities and colleges we see students with cellphones and laptops. Creating CALL software that is accessible on a variety of platforms and devices may not make the technology more convenient for teachers, but it will eliminate a possible inconvenience. A subset of accessibility is availability. High costs of language learning software programs are inconvenient for teachers, students, and program administrations. Even though we conducted this study in a university with access to educational pricing, none of the participants indicated that they used language-learning-specific software, which was purchased either personally or by their departments. One exception to this would be the textbooks that come with accompanying software or media. Bundling language learning software with

textbooks may be a way for developers to make it more convenient for their users to use in their teaching. Additionally, software developed in-house is more accessible provided that the resources are available to have such software developed.

Language teachers. The implications for language teachers compliment those of teacher educators and administrators.

Consider learning outcomes first. Because pedagogy was such a predominant theme in this multiple case study, teachers who wish to become expert technology users should evaluate their own criteria when choosing to use CALL. Making pedagogical concerns the first priority will help shape and focus their technology use. It is essential that all teachers consider the importance of student learning outcomes when considering the use of technology in the classroom.

Practice good pedagogy. Teachers who base instruction in sound pedagogy are more likely to use technology more effectively. As such, teachers who wish to be better users of technology in the classroom should start by reinforcing the effective teaching practices they currently use and be willing to improve their quality of teaching. Observing other teachers, being observed by others, and video recording their own teaching are possible ways to evaluate, reflect upon, and improve teaching.

Use CALL resources and activities. As mentioned earlier, one sub-theme of convenience was familiarity and this perhaps provides us with the key implication for language teachers: experience using CALL may be a key factor in the use of CALL. If expert technology users select CALL materials and activities based on their familiarity with the technology, teachers who wish to use technology should become more experienced in its use. Meskill, Mossop, DiAngelo, and Pasquale (2002) also reported that teachers with high expertise in CALL

felt more confident in its usage. Wetzel, Zambo, and Ryan (2007) also reported the effect of experience on technology use. Teachers with experience in CALL become familiar with CALL. In essence, teachers who wish to use technology more effectively should seek to have more experience in using technology.

Suggestions for Future Research

Despite the insights provided in this study, it also generated several questions that lead to future research in CALL. First, if these criteria— pedagogy, convenience, and authenticity—are used by expert technology users, what criteria do novice technology users consider? In a similar yet divergent topic, is there a correlation or other relationship between teachers who are expert technology users and those that practice good pedagogy in their teaching? Is the practice of good pedagogy a predictor for good technology use? Second, what criteria do program administrators consider when choosing to adopt software or hardware? How do these differ from the criteria teachers consider? Third, what criteria do students consider when choosing to use technology to aid in language acquisition? The results from this study in addition to future research regarding the criteria used by administrators and students will further help inform teachers, learners, administrators, and developers in the creation and use of CALL.

Another interesting topic that this study revealed concerns the use of materials that were especially designed for language learning. Three participants used materials that were in varying degrees created for the purpose of language learning. One participant used in-house developed software geared at language learning. Another participant used materials that were part of a language textbook. A third participant used authentic materials that were incorporated into a website used to teach a foreign language. Perhaps the predominant question is *who is using*

what? How many language teachers are actually using CALL- specific materials? What CALL specific materials are language teachers using?

Factors inhibiting or promoting the use of technology in all institutions should be researched. Furthermore, research should be done to investigate the role of administrators in teachers' use of technology as well as institutional factors that may affect teachers' use of technology.

Article References

- Baniabdelrahman, A. A., & Bataineh, R. F. (2007). An exploratory study of Jordanian EFL students' perceptions of their use of the Internet. *IATEFL Poland, Teaching English with Technology, 7* (3).
- Bordonaro, K. (2003). Perceptions of technology and manifestations of language learner autonomy. *CALL-EJ Online, 5*(1), 1–21.
- Brantmeier, C. (2003). Technology and second language reading at the university level: informed instructors' perceptions. *Reading, 3*(3).
- BYU. (2012). Languages at BYU. Retrieved 03/12/2012, from <http://yfacts.byu.edu/viewarticle.aspx?id=139>.
- Cook, K., McGhee, J., & Lonsdale, D. (2011). Elicited imitation for prediction of OPI test scores. *Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications, IUNLPBEA '11* (pp. 30–37). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from <http://dl.acm.org/citation.cfm?id=2043132.2043136>
- Egbert, J., Paulus, T. M., & Nakamichi, Y. (2002). The impact of CALL instruction on classroom computer use: A foundation for rethinking technology in teacher education. *Language, Learning & Technology, 6*(3).
- Hatch, J. A. (2002). *Doing qualitative research in education settings*. Albany, NY: State University of New York Press.
- Kessler, G. (2007). Formal and informal CALL preparation and teacher attitude toward technology. *Computer Assisted Language Learning, 20*(2), 173–188.

- Kessler, G., & Plakans, L. (2008). Does teachers' confidence with CALL equal innovative and integrated use? *Computer Assisted Language Learning*, 21(3), 269–282.
- Kim, H. K. (2008). Beyond motivation: ESL/EFL teachers' perceptions of the role of computers. *CALICO Journal*, 25(2), 241–259.
- Kung, S. C., & Chuo, T. W. (2002). Students' perceptions of English learning through ESL/EFL websites. *TESL-EJ*, 6(1), 1–14.
- Lee, L. (2005). Using web-based instruction to promote active learning: Learners' perspectives. *CALICO Journal*, 23(1), 139.
- Meskill, C., Mossop, J., DiAngelo, S., & Pasquale, R. K. (2002). Expert and novice teachers talking technology: Precepts, concepts, and misconcepts. *Language, Learning & Technology*, 6(3), 46–57.
- Moustakas, C. E. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage Publications, Inc.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Sage Publications, Inc.
- Peters, M., Weinberg, A., & Sarma, N. (2009). To like or not to like! Student perceptions of technological activities for learning French as a second language at five Canadian universities. *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 65(5), 869–896.
- Reid, K., Flowers, P., & Larkin, M. (2005). Exploring lived experience. *The Psychologist*, 18(1), 20–23.
- Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method and research*. London: Sage Publications Ltd.

- Son, J. B. (2007). Learner experiences in web-based language learning. *Computer Assisted Language Learning*, 20(1), 21–36.
- Stake, R. E. (2010). *Qualitative Research: Studying How Things Work* (1st ed.). New York, NY: The Guilford Press.
- Stepp-Greany, J. (2002). Student perceptions on language learning in a technological environment: Implications for the new millennium. *Language, Learning & Technology*, 6(1).
- Van Manen, M. (1997). *Researching lived experience: Human science for an action sensitive pedagogy*. Ontario: Althouse Press.
- Wetzel, K., Zambo, R., & Ryan, J. (n.d.). Contrasts in classroom technology use between beginning and experienced teachers. *Learning*, 3(1), 15–27.

OVERALL DISSCUSSION

The overall theme of this dissertation revolves around evaluation practices in computer-assisted language learning. With a background in language acquisition, experience and interest in CALL, and recent study in the field of educational evaluation, I quickly saw gaps in CALL evaluation. At the onset, I noticed a few trends in CALL publications.

Trends in CALL Literature

First, many of the publications in CALL were types of evaluations. In many instances, researchers developed a CALL product and implemented standard research practices to evaluate the author-generated CALL material. While many of these were excellent examples of evaluation, they still lacked essential evaluation practices that would strengthen the evaluation such as considering various stakeholders, and the selection of evaluation criteria.

Second, CALL scholars have proposed various evaluation frameworks based in part on personal experience and expertise, and also in generally accepted research practices. While these frameworks have benefited CALL evaluation for years and are in many cases the foundation of software reviews in peer-reviewed journals, they failed to tap into the wealth of knowledge and experience of formal evaluators, their proposed frameworks, and generally accepted evaluation practices.

Third, CALL research lacked studies that explored the informal evaluation practices that practitioners regularly participate in. Teachers evaluate daily when they select materials and activities for use in their classrooms. One immediate question that arose focused on the criteria these teachers consider when selecting CALL materials to enhance language learning. With a better understanding of this criteria program administrators, teacher educators, software

developers, and teachers may be able to facilitate the use of, use, and develop, CALL technologies.

Conclusion

In conclusion, formal CALL evaluation would benefit from the borrowing and implementation of formal evaluation principles and tasks and there is still more research to be done in exploring the informal practices of language teachers.

Formal CALL evaluation. Institutions and journals can increase the quality and effectiveness of CALL evaluation by following formal evaluation principles and tasks. Evaluations that consider the values of stakeholders may make these evaluations more transferable. Considering the nine evaluation tasks presented, popular CALL evaluation frameworks are not without merit, but may benefit from the inclusion of formal evaluation principles and tasks. Future research regarding the use of CALL evaluations and evaluation standards may also help enhance formal evaluation in CALL.

Informal CALL evaluation. With a better understanding of the criteria expert technology users consider, developers and administrators can make important considerations in the development and adoption of CALL, respectively. Developers can focus on creating CALL specific technology solutions that focus on pedagogy, convenience, and authenticity. Likewise, program administrators can be mindful of these three criteria in making decisions regarding institutional adoption of CALL products.

Additionally, future CALL research may explore the criteria novice technology users avail themselves of. This research only represents a segment of the population of language teachers. The qualitative nature of the study provided an in depth look at expert technology

users, but quantitative and mixed-method approaches may also serve to answer similar questions regarding a larger gamut of language teachers who use or do not use CALL.

DISSERTATION REFERENCES

- Beatty, K. (2010). *Teaching and researching: Computer-assisted language learning* (2nd Revised ed.). London, England: Pearson Education Limited.
- Burston, J. (2003). Software selection: A primer on sources and evaluation. *CALICO Journal*, 21(1), 29–40.
- Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge, England: Cambridge University Press.
- Chapelle, C. (2007). Challenges in evaluation of innovation: Observations from technology research. *Innovation in Language Learning and Teaching*, 1(1), 30–45.
- Chapelle, C. (2010). The spread of computer-assisted language learning. *Language Teaching*, 43(1), 66–74.
- Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage Publications, Inc.
- Egbert, J. L. (2005). Conducting research on CALL. In J. L. Egbert & G. M. Petrie (Eds.), *CALL research perspectives* (pp. 3–8). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Garrett, N. (2009). Computer-assisted language learning trends and issues revisited: Integrating innovation. *The Modern Language Journal*, 93(focus issue), 719–740.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass Inc.
- Hubbard, P. (1987). Language teaching approaches, the evaluation of CALL software and design implications. *Modern Media in Foreign Language Education: Theory and Implementation* (pp. 227–254). Lincolnwood, IL: National Textbook Company.

- Hubbard, P. (1988). An integrated framework for CALL courseware evaluation. *CALICO Journal*, 6(2), 51–72.
- Hubbard, P. (1996). Elements of CALL methodology: Development, evaluation, and implementation. In M. Pennington (Ed.), *The Power of CALL* (pp. 15–32). Houston, TX: Athelstan.
- Hubbard, P. (2006). Evaluating CALL software. In L. Ducate & N. Arnold (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching* (pp. 313–318). San Marcos, TX: CALICO.
- Hubbard, P. (2011). Evaluation of courseware and websites. In L. Ducate & N. Arnold (Eds.), *Present and Future Promises of CALL: From Theory and Research to New Directions in Foreign Language Teaching*, Second Edition. San Marcos, TX: CALICO.
- Levy, M., & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer-assisted language learning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Lincoln, Y. S. (2003). Constructivist knowing, participatory ethics and responsive evaluation: A model for the 21st century. *International Handbook of Educational Evaluation*, 69–78.
- Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation*, 22, 329–336.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage Publications, Inc.
- Patton, M. Q. (2003). Utilization-focused evaluation. *International Handbook of Educational Evaluation*, 9(1), 223–244.

- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Reeder, K., Heift, T., Roche, J., Tabyanian, S., Schlickau, S., & Golz, P. (2004). Toward a theory of e/valuation for second language learning media. In S. Fotos & C. Browne (Eds.), *New perspectives on CALL for second language classrooms* (pp. 255–278). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Reeves, T. C., & Hedberg, J. G. (2003). *Interactive learning systems evaluation*. Englewood Cliffs, NJ: Educational Technology Publications, Inc.
- Scriven, M. (1974). Evaluation perspectives and procedures. In W. J. Popham (Ed.), *Evaluation in Education*. Berkeley, CA: McCutchan Publishing Corporation.
- Scriven, M. (1990). The evaluation of hardware and software. *Studies in Educational Evaluation*, 16(1), 3–40.
- Scriven, M. (2003). Evaluation theory and metatheory. *International Handbook of Educational Evaluation*, 15.
- Stake, R. E. (2003). Responsive evaluation. In E. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (Vol. 9, pp. 63–68). Springer International Handbooks of Education.
- Stake, R. E. (2004). *Standards-based and responsive evaluation*. Thousand Oaks, CA: Sage Publications, Inc.
- Stake, R. E. (2010). *Qualitative research: Studying how things work* (1st ed.). New York, NY: The Guilford Press.
- Stake, R. E., & Schwandt, T. A. (2006). On discerning quality in evaluation. *The Sage Handbook of Evaluation*, 404–418.

- Stufflebeam, D. L. (2003a). Professional standards and principles of evaluations. *International Handbook of Educational Evaluation*.
- Stufflebeam, D. L. (2003b). The CIPP model for evaluation. *International Handbook of Educational Evaluation*, 9(1), 31–62.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco, CA: Jossey-Bass Inc Pub.
- Susser, B. (2001). A defense of checklists for courseware evaluation. *ReCALL*, 13(02), 261–276.
- Tomlinson, B. (2007). *Materials development in language teaching*. Ernst Klett Sprachen.
- Tomlinson, B. (2003). Materials evaluation. In B. Tomlinson (Ed.), *Developing materials for language teaching* (pp. 15–36). Trowbridge, Wiltshire: Cromwell Press.
- Van Manen, M. (1997). *Researching lived experience: Human science for an action sensitive pedagogy*. Ontario: Althouse Press.
- Villada, E. G. (2009). CALL evaluation for early foreign language learning: A review of the literature and a framework for evaluation. *CALICO Journal*, 26(2), 363-389.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.