

Article

Methodological Diversity in Language Assessment Research: The Role of Mixed Methods in Classroom-Based Language Assessment Studies

Rika Tsushima, PhD (Cand)
Department of Integrated Studies in Education
McGill University
Montréal, Canada

© 2015 Tsushima. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-NonCommercial-ShareAlike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and if transformed, the resulting work is redistributed under the same or similar license to this one.

Abstract

Although an epistemic change often is labeled as a *shift*, some researchers representing the social sciences consider this paradigm shift as a paradigm expansion (e.g., Pollack, 2007) because epistemological and methodological diversity allows researchers to address a wider range of questions. The third paradigm, mixed methods (MM) research, is claimed to provide a more holistic picture of a research problem by combining two different data sources—quantitative and qualitative—in a single study (Creswell & Plano Clark, 2007; Greene, 2007; Teddlie & Tashakkori, 2009). This article discusses how mixed methods research approaches have been used to enrich the results and to enhance the rigor of classroom-based language assessment investigations, drawing on both the language testing and assessment (hereafter, *language assessment*) and classroom assessment literature in second language education. The article opens with a brief overview of the methodological evolution in language assessment research. Then, focusing on MM research studies that investigated various facets of classroom-based language assessment (CBLA) practices, the main part of this article outlines a proposal of MM methodology as an appropriate methodology for research on CBLA, particularly in a context where a new form of assessment is implemented. The article closes with a discussion of challenges that MM researchers might face, and a proposal of MM research methodology as an appropriate research approach for CBLA scholars, especially in settings where both the validation of assessment and the explanation of the phenomenon are required.

Keywords: classroom-based assessment, formative assessment, language assessment, mixed methods research, second language education

Research

Language testing and assessment (hereafter, language assessment) is a subfield in applied linguistics that utilizes an array of qualitative and quantitative tools in order to evaluate students' use of language. Over the last decade, language assessment research has been expanding its scope of inquiry. In particular, owing to the increased attention to classroom language assessment accompanied by educational policy reforms in many countries, researchers have become more involved in issues of language assessment in school and classroom settings (Bachman, 2007). Owing to newly emerging phenomena around the complex nature of assessing learners' language development in classrooms, language assessment researchers have been seeking for a research methodology that allows them to address their research questions more effectively. In many cases, when investigating the implementation of a new form of assessment in language classrooms, researchers need to address research questions that intrinsically require both quantitative and qualitative data. In a state-of-the-art article, Bachman (2000) described how in conjunction with the theoretical development and expansion of this field of study, the methodological approaches employed in the field have become increasingly diverse in the past two decades. Indeed, Cheng and Fox (2013), examining recent doctoral dissertations in language assessment, identified the increasing use of multiple methods approaches in comparison with single method studies.

Historically speaking, researchers in language assessment were trained in and equipped with quantitative research methods. However, acknowledging that the contribution qualitative methodology can make, the research community also welcomed such an approach. At present, both quantitative and qualitative research methodologies are largely used in research in this field; furthermore, some methodological approaches that were either nonexistent or hardly used 30 years ago have become regular, mainstream tools for language assessment research and practice (Bachman, 2007). Although it has been known that many language assessment researchers work within the postpositivist paradigm that underlies most quantitative research, the number of researchers working within other paradigms and the volume of their research has been growing rapidly. Paradigms, defined by Kuhn (1970), are an "entire constellation of beliefs, values, techniques shared by members of a given scientific community" (p. 75). Following this concept, researchers who adopt the qualitative methodology fall under the interpretivist paradigm. Then, researchers in the "third research community" (Teddlie & Tashakkori, 2009, p. 4), in which both quantitative and qualitative data are employed often, are considered as members of the pragmatist paradigm.

Representing the third community of research, according to Creswell, Plano Clark, Gutmann, and Hanson (2003), mixed methods (MM) research is defined as:

The collection or analysis of both quantitative and qualitative data in a single study in which the data are collected concurrently or sequentially, are given a priority, and involve the integration of the data at one or more stages in scientific research. (p. 212)

By combining two different data sources within one study, the MM research framework is claimed to provide a more holistic picture of a research problem (Creswell & Plano Clark, 2007; Greene, 2007; Teddlie & Tashakkori, 2009). Based on their argument, this article contains a discussion of comprehensiveness of incorporating qualitative data into mixed methods research in order to address a wider range of questions than quantitative methods alone would provide. Furthermore, it focuses on how such approaches have been used to enrich the results and to enhance the rigor of classroom-based language assessment (CBLA) investigations, drawing on both the language assessment and classroom assessment literature in second language education. Followed by a brief overview of the methodological evolution in language assessment research, in this article, I provide examples of MM studies that investigated various facets of CBLA

practices to portray how researchers can advance the scholarly knowledge in the area by taking advantage of this methodology. This article closes with a discussion of challenges that MM researchers might face, and a proposal of MM methodology as an appropriate research approach for CBLA inquiry, particularly in a context where both the validation of assessment and the explanation of the phenomenon are required.

A Brief Overview of Methodological Expansion in Language Assessment

The Historical Orientation toward Quantitative Research

In like manner with other fields of social science, in many disciplines of applied linguistics, quantitative research has been actively employed for studies that test a theoretical hypothesis, rather than to extract new knowledge, posed by a researcher (Jamieson & McDonough, 2012). The methodological tools for data collection in applied linguistics include tests, interviews, naturalistic data, surveys, verbal reports, prompted production, prompted responses, and grammar analysis techniques (Gass & Mackey, 2007), and these means of data collection appear in CBLA research as well. Language assessment has a tradition of using psychometric procedures and a strong orientation toward quantitative research methods in the quest for test score validity (Bachman, 2007; Purpura, 2011). Such a trend has apparently been reflected in the research agenda of this scholarly community. For example, an abstract analysis conducted by Hamp-Lyons and Lynch (1998) using data from the Language Testing Research Colloquium from 1979 to 1994, revealed that studies on validity in language assessment reflected predominantly postpositivist thinking at the time with a strong dependence on psychometric methods to support claims of validity. Similarly, Chalhoub-Deville and Deville (2008) conducted an analysis of articles in the two major language assessment journals (i.e., *Language Assessment* and *Language Assessment Quarterly*) from their beginnings to 2005, and concluded that quantitative research approaches were dominantly used in the publications.

This strong orientation toward quantitative research methodology stems from the dominant research interest in language assessment revolving around the development of score-valid and score-reliable large-scale standardized tests (Chalhoub-Deville & Deville, 2008; Hamp-Lyons & Lynch, 1998). Hamp-Lyons (2000), describing the developments in the philosophy and epistemology of language testing, stated that the epistemology of early language tests was shaped by a faith in psychometrics and positivism due to a strong influence from intelligence tests. Back in the mid-1960s, language testing research was dominated largely by the hypothesis that language proficiency consisted of a single unitary trait, and that a quantitative, statistical research methodology was the most appropriate for understanding it (Bachman, 2000, 2007; Stoyhoff, 2012). Moreover, there are other factors that account for the popularity of quantitative research methodology. For instance, language testing validation research places a fundamental emphasis on the generalizability of results and the appropriateness of inferences based on observed learner performances (McNamara, 2003), and the results of language tests are still most frequently reported quantitatively (Bachman, 2007). Purpura (2011) has stated that as validation frameworks have become more sophisticated, quantitative research methods used for validation related research have been diverse and increased (e.g., generalizability theory, item-response theory, structural equation modeling). His observation of current research methodologies used in the field reflects the view that quantitative research methodology remains the commonly utilized solution for understanding some issues in particular.

The Introduction of Qualitative Research

Since the late 1990s, however, due to epistemological changes in second language acquisition research and the increasing awareness that any language assessment cannot be separated from the context, culture, and values of where it was created and used (McNamara, 2001), hermeneutic and qualitative approaches have been adopted to gain deeper insights into issues concerning

designing, administering, and interpreting language assessment (Banerjee & Luoma, 1997; Chalhoub-Deville & Deville, 2008). This reflects the general methodological development in language education, which can be characterized regarding a philosophical position shift from a postpositivist worldview toward interpretive (Mezirow, 1996; Moss, 1996) or interpretivist (Lynch & Shaw, 2005) approaches. That is, the core belief of the field moved from the behaviouristic, transmission model of teaching to constructivist, learner-centered pedagogy that emphasizes the importance of contextualization. Adopting Sociocultural Theory (Vygotsky, 1978), the interpretive/interpretivist epistemic orientation stems from the belief that multifaceted psychological processes can only be understood within the sociocultural context (Wajda, 2011). In this paradigm, as opposed to the postpositivist perspective, learning and knowledge are generated through meaning construction in their social world rather than objective reflections of an external world (Piaget, 1975; Rogoff, 1999; Vygotsky, 1978). From the perspective of this epistemological position, language assessment researchers started employing qualitative research methodology so as to address issues that require an analysis of content-related variables (e.g., the effects on test performance of test takers' characteristics, the strategies used to respond to assessment tasks, the nature of the discourse produced in the process of language assessments). Furthermore, sociocultural views of language assessment have invited more diverse tools to measure and to enhance classroom learning. Even in the 1990s qualitative research approaches were used in a complementary manner with or without being labeled as MM research (e.g., Anderson, Bachman, Cohen, & Perkins, 1991; Clapham, 1996; Weigle, 1994), which was a sign of further methodological expansion in the field. In this sense, postpositivist scholars who were in need of incorporating qualitative research frameworks into their research contributed to shaping the mixed methods discourse in this field (Denzin, 2010).

The Advent of the Hybrid Methodology

It was around the beginning of the twenty-first century when language assessment researchers started gradually framing their methodologies as representing MM research (Turner, 2013). Adopting a pragmatic worldview as the philosophical basis in which the research question is the decisive factor for selecting the research methodology—not the paradigm from which the method derives (Teddlie & Tashakkori, 2010)—researchers have chosen to utilize MM in their studies to answer questions that could not be answered monolithically (Johnson & Onwuegbuzie, 2004).

Turner (2013), examining MM studies published in the three major journals in language assessment (i.e., *Assessing Writing*, *Language Testing*, and *Language Assessment Quarterly*), has reported that, by and large, the evolution of MM research seems analogous to the general trajectory of methodology in the social and behavioral sciences. Her analysis demonstrates that up to approximately 2003, the term MM research was rarely mentioned in the methodology section of published articles despite the fact that the research approaches and reports of data interpretation suggested the integration of qualitative and quantitative data in a single study. Aside from Turner's (2013) work, this scholarly community, however, has not developed a specific body of literature that argues for the usefulness and implications of MM for the field. Consequently, even though many disciplines in the health, social, and behavioral sciences have been developing frameworks for conducting MM research (e.g., Collins, Onwuegbuzie, & Sutton, 2006 in special education; Creswell, Fetters, & Ivankova, 2004 in primary care; Rocco, Bliss, Gallagher, Pérez, & Prado, 2003 in organization systems), MM has not been adopted yet to a similar degree in language assessment. Although further discussion of the methodology is necessary, in the past several years, language assessment researchers have been increasingly finding that the relevance and significance of research can be enhanced by this new methodological approach (Bachman, 2007; Turner, 2013). Accordingly, authors of language assessment studies published in recent peer-reviewed journals are more likely to label their own methodological approaches as MM research (e.g., Neumann, 2010; Tan, 2011), and the number

of emerging scholars who employed MM approaches for their doctoral studies rather than the other two methodologies seems to be growing in this field (Cheng & Fox, 2013). In concurrence with Johnson and Onwuegbuzie's (2004) argument that each of the three paradigms (quantitative, qualitative, MM) can be superior under different circumstances, when a monolithic methodology is not enough to address research questions and to achieve research objectives, MM approaches can serve researchers to generate important and innovative insights in the field.

The Methodological Diversity in Classroom-based Language Assessment Research

Aligned with the current trend of educational reforms at both the domestic and international levels, there has been a major shift from conventional means of one-size-fits-all type educational measurement to more context-specific assessment formats to examine and to enhance students' language learning in the classroom (Carless, 2006; Purpura, 2004; Turner, 2012; Tuner & Purpura, 2013). CBLA, in this article, refers to activities that are specifically designed and used internally for the purpose of grading, evaluating, testing, ranking, and/or gleaning information of students' language learning development, as opposed to large-scale standardized examinations. To date, language assessment experts have been developing a growing interest, body of research, and practice in issues related to CBLA as well as appropriate research methods for such studies.

The Qualitative Perspectives in CBLA Research

As discussed earlier, although quantitative research methodology has been used predominantly in language assessment, researchers acknowledge that due to the nature of classroom dynamics, there are numerous factors outside of the research focus that might be affecting learners' behaviour. In this regard, some elements that are beyond the researcher's knowledge require a different approach, such as the use of qualitative research methods. Classroom researchers in second language education have been promoting and employing qualitative approaches, acknowledging the value of "thick description" (Geertz, 1973) to conceptualize and to capture classroom dynamics (Bailey & Nunan, 1996). Due to the complex nature of classroom assessment as a process of collecting and analyzing the evidence of student learning outcomes, the beliefs and perspectives of the assessors—classroom teachers—often are used as essential data to construct a more comprehensive argument. Thus, many CBLA researchers consider the necessity of approaching this issue from the perspective of an interpretive, qualitative research methodology.

Mohan (2012), discussing how qualitative research methods have been contributing to the understanding of the complex nature of language assessment practices, stated that the community needs a methodology that "incorporates a linguistic model of language that relates wording and meaning" (p. 762). Furthermore, Lumley and Brown (2005), analyzing publications in language assessment, found that three qualitative research methods—discourse analysis, introspection, and ethnographic methods—were used mainly to discuss the problems in assessment practices. For example, Scarino (2013) used introspection and Ishihara (2009) used ethnographic methods to investigate influences of implementation of formative assessment on learning. Simply speaking, formative evaluation serves as a tool to glean information from students to make decisions about what to teach next to advance their learning and is meant to support students' further learning, whereas summative assessment provides certification of achievement (Scriven, 1967). In Ishihara's (2009) ethnographic case study, for instance, she analyzed Japanese undergraduate ESL students' written reflections to investigate the effectiveness of teacher feedback on their pragmatic competence. In order to understand the effects of the innovative assessment practice in her classroom, Ishihara took an interpretative approach. Likewise, many CBLA researchers have utilized qualitative research methods in order to provide a contextualized perspective to discuss issues concerning language assessment practices in different classroom contexts that require a thick description of complex phenomena (Geertz, 1973).

In Pursuit of a More Appropriate Methodology for CBLA Research

Qualitative methods, however, are not always effective or practical in investigating such complex phenomena at the broader, regional or national level involving a large number of stakeholders. Although systematically documented qualitative data help other researchers to analyze the cases and the results, and to make decisions about aspects that might or might not be applicable to other populations and contexts (Duff, 2006), knowledge produced through qualitative research methods cannot be used to make quantitative predictions or to test hypotheses and theories with large participant pools, which are historically essential aspects of language assessment research. Thus, as a result of seeking a pertinent methodology, the use of both quantitative and qualitative research approaches in a single study (i.e., MM research) has been embraced in some CBLA research (e.g., Colby-Kelly & Turner, 2007; Gunning, 2011; Jeong, 2013; Lee & Coniam, 2013). The introduction of the third research paradigm (i.e., MM research) in CBLA research has provided more flexibility to investigate emerging issues that call for different approaches from the two pre-existing research methodologies. Mixed methods, in this respect, is a merged and expanded form of the two research methodologies for this particular research community.

The Role of a MM Approach in Formative Assessment Research

Stimulated by the current movement toward more effective classroom teaching that facilitates better learning outcomes, formative assessment has been attracting increasing attention from language assessment researchers. To date, a substantial number of researchers have attempted to align formative assessment with contemporary psychological theories of learning (e.g., Black, Harrison, Lee, Marshall, & Wiliam, 2003; Black & Wiliam, 1998; Boud, 1995; Gipps, 1995) in order to explore appropriate ways to embed language assessment into daily instruction for the purposes of learning enhancement. Examining MM research approaches used in such a context might elucidate how this dual approach has been contributing to the research. As mentioned earlier, however, owing to the quiet emergence of MM research into the field, not all CBLA studies that involved the use of a MM research approach provide clear and detailed information in their methodology sections.

For example, Guardado and Shi (2007) apparently employed a MM research design in their study on ESL students' experiences of peer feedback on their essays at a Canadian university. Even though not explicitly stated in their article, their study can be labeled as MM research. In this study, the quantitative data were collected and analyzed to examine the efficacy of formative peer assessment implemented in the course, and the qualitative data were necessarily used to explain why the phenomena happened in this setting. The findings from the two data sets were compared, contrasted, or related in order to "seek elaboration, enhancement, illustration and clarification of the results from one method with the results from the other method" (Creswell & Plano Clark, 2011, p. 62).

An exemplar that contains a clearer presentation of the methodological standpoint is found in the more recent literature. In Ghoorchaie, Tavakoli, and Ansari's (2010) research on the impact of portfolio assessment as a process-oriented assessment mechanism on Iranian English as a foreign language (EFL) students' writing ability, the primary data collection phase was organized as a quasi-experimental study, in which the experimental group received the treatment of formative assessment (i.e., portfolio assessment), whereas the control group underwent the conventional nonformative assessment. The results from the quantitative data analysis revealed that the portfolio assessment group outperformed the control group overall. Additionally, semi-structured interviews were conducted to "understand students' perceptions about the impact captured via quantitative results" (p. 38). These researchers concluded that their hypothesis regarding the prediction of the formative potential of portfolio assessment in EFL classes was confirmed and supported by integrating the two types of data sets.

In their research on the implementation of formative assessment for EFL writing within an examination-driven system in Hong Kong, Lee and Coniam (2013) presented a case study research design created as part of a larger study. They used a MM research approach because “the multiple sources of the data in the study provide triangulation to validate the analysis and shed light on the research questions” (p. 39). Drawing on data that included questionnaires, teacher and student interviews, pretests and posttests, and lesson observations, the researchers investigated the effects of formative assessment on students’ motivation and writing performance, as well as the factors that might facilitate or inhibit its uptake. The results of both quantitative and qualitative data indicated that this assessment format had positive influences on students’ writing performance. However, the student questionnaires revealed no statistically significant difference between students’ motivation before and after the treatment. These contradictory findings were explained further via the interviews on 12 students, and some tension between innovative and traditional assessment practices also were captured through the integration of the data. In other words, the integration technique helped the researchers not only test the effects of the implemented assessment practice but also describe the complexity of interaction between teaching and learning. In these studies, the role of qualitative components was mainly supplemental to describe a phenomenon or to evaluate the effect of introduced assessment practice. Due to the current demands for CBLA studies, it is predictable that the role of MM research will be more apparent and diverse, and so will the qualitative components.

The Further Expansion of MM Research Approaches through CBLA Research

It is noteworthy that many recent CBLA researchers seem to have employed multiple data sets in their MM research studies (e.g., Colby-Kelly & Turner, 2007; Guardado & Shi, 2007) and the design of their data collection and analysis procedures were very diverse and unique in each study. Turner (2013), in a state-of-the-art article on MM research studies in language assessment, presented the six major families of MM research designs defined by Creswell and Plano Clark (2007, 2011). In conjunction with the current language education research becoming more diverse and complex, researchers have started customizing research designs that fit their unique context. In this methodology, researchers need to consider not only types of data, but also the order of data collection, how the data sets will be analyzed, and how they are integrated. Methodology experts have introduced different types of MM research designs (e.g., Creswell & Plano Clark, 2007; Creswell et al., 2003; Leech & Onwuegbuzie, 2009), but it might be more accurate to state that each of the research designs can be developed further into numerous research designs for the purposes of obtaining answers to our complex research questions.

Fitting the Context: Adopting a MM Research Design

An example of MM research involving the adoption of an established MM design is the research design employed by Tsushima (2011), which illustrates how multiple sources of data can produce the more complete knowledge that is necessary to depict the complex nature of teaching and assessment practices in language classrooms. The study involved an investigation of the perspectives of Japanese teachers of English as a foreign language (EFL) at the secondary school level regarding the alignment of their teaching and testing practices with their speaking-focused course objectives. In this research context, due to the strong influence of university entrance examinations nationally, it is often argued that, in practice, speaking-focused courses do not focus on communication-oriented learning activities but on analytical, grammar-oriented exercises designed to prepare students for the high-stakes examinations (Kikuchi & Browne, 2009). To investigate the reality of such courses, the following research questions were formulated:

1. How are teaching and testing in speaking-focused courses of English in Japanese secondary schools related to each other when the educational policy and university entrance examinations are not in alignment?
2. What are the nature and the scope of influences from university entrance examinations on teaching and testing in the speaking-focused courses, and how do teachers perceive such influences?

A MM Study Design with Multiple Stages

In order to address the research questions, Tsushima (2011) modified Creswell and Plano Clark's (2007) explanatory sequential design (see Figure 1). Generally, the explanatory sequential design has two phases; that is, at first the investigator collects and analyzes quantitative data and, then, qualitative data are collected and analyzed to obtain explanations of the quantitative results (Creswell & Plano Clark, 2007, 2011). In this study, however, the modified MM research model comprised three phases with specific goals that required four data sets:

Phase 1 Broader view: A teacher survey was administered to obtain broader views of the issue from the target population. The researcher required the data to prepare the subsequent data collection instruments, and to decide types of schools as data collection sites.

Phase 2 Deeper View: The two qualitative data sets, which consisted of classroom observations and term examination samples, were included to obtain a deeper understanding of the actual classroom organization, teaching, and testing. Another goal of this phase was to formulate interview questions to identify reasons for the gap between teaching and testing practices.

Phase 3 Specific View: The researcher conducted guided interviews to obtain teachers' insights into the phenomenon of classroom teaching and assessment. The interview data were analyzed, compared, and integrated with the three data sets from the previous two phases.

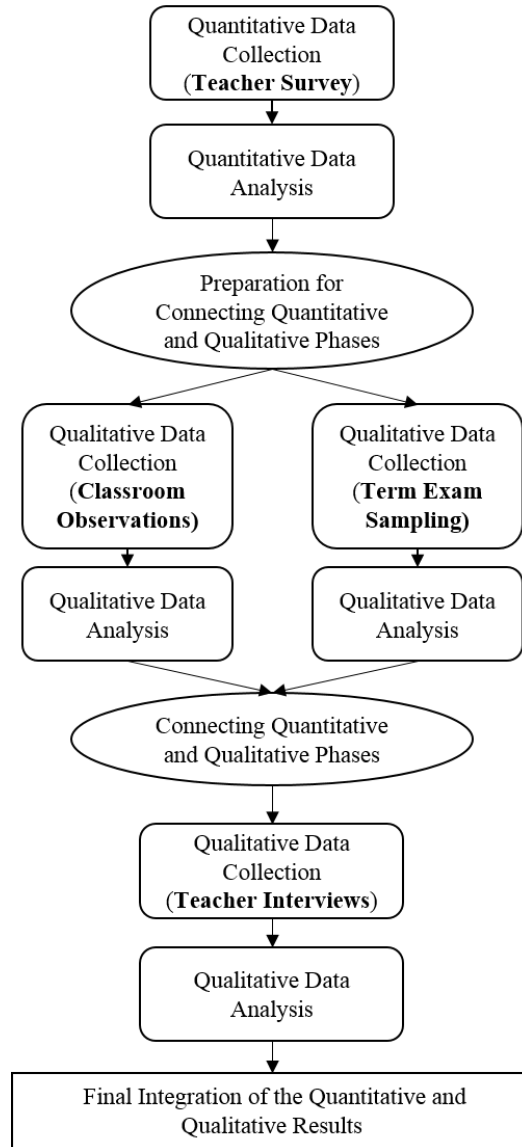


Figure 1. An explanatory sequential mixed methods research design with multiple phases for a study on teaching and testing practices in EFL speaking courses.

Designing Instruments and Analyzing the Data for Knowledge Integration

Each of the three stages was necessary in order to obtain answers to the broader and more complete range of research questions, and to acknowledge the multifaceted nature of the phenomenon in classrooms. The three stages were designed to garner information from different angles and to integrate the whole knowledge to address the research questions in a cohesive manner. The first set of data indicated that although teachers ($N = 87$) were adopting communication-focus instruction, the assessment of the courses remained the traditional grammar-focused form. Moreover, statistically, this tendency was more evident in academically oriented schools, where the majority of students would go on to university, than vocationally oriented schools.

Informed by the quantitative data, the researcher designed instruments to collect qualitative data from five teachers working in five different schools. The two qualitative data sets agreed with the quantitative results of the phenomenon of classroom teaching and assessment not being congruent

with each other or with the course objectives. At this point, the researcher obtained information on the classroom practice; however, still missing was the explanation for the mismatched classroom practices. Thematic analyses of the guided interviews with nine teachers in the five schools suggested that because of the strong pressure to get their students into prestigious universities, the teachers in academically oriented schools considered the high-stakes examinations as a more influential factor than was the educational policy; therefore, term examinations were used as convenient tools to align with the content of university entrance examinations even though they were aware of the mismatch between their teaching and testing practices. In this MM research study, the qualitative components played a dual role that is a unique characteristic of the MM research approach: (a) a supporting role to the quantitative analysis; and (b) a key role in obtaining a more comprehensive and unequivocal explanation for the phenomenon under investigation. As seen in this example, driven by the research questions, the researcher chose a MM research approach and tailor-made the best fitting research design. Due to the intrinsic value of language assessment research in a classroom context, researchers are likely to develop research questions that cannot be answered without multiple data sets. Therefore, with its expandability and customizability, a MM methodology enables CBLA researchers to explore and to capture their target phenomenon in a holistic manner, which makes the research findings more reliable and valid. At the same time, nonetheless, it should be noted that MM studies in the literature tend to expose the intricacy and difficulty of such research at different levels. Although a MM research approach can maximize the possibilities of obtaining a better understanding of the target issue, it is of a great importance to consider some of the challenges that MM researchers tend to face.

General Challenges of Conducting MM Research in Classrooms

The goal of MM research is not to replace quantitative or qualitative research approaches, but to draw synergistically from their strengths and to minimize their limitations (Johnson & Onwuegbuzie, 2004); thus, MM researchers need to be open to those limitations and to learn the best possible means to reduce weaknesses in their empirical work. Teddlie and Tashakkori (2003) addressed six key issues associated with the use of MM research methodology: (a) the nomenclature and basic definitions used in mixed methods research; (b) the utility of mixed methods research; (c) the paradigmatic underpinning for mixed methods research; (d) design issues in mixed methods research; (e) issues in making inferences in mixed methods research; and (f) the logistics of conducting mixed methods research. Although the discourse on the issues has been significantly developed by research methodology experts (e.g., Creswell & Plano Clark, 2007; Greene, 2007; Johnson & Onwuegbuzie, 2004), some of these issues are still unsolved today as inherent characteristics of the methodology.

Theoretical Challenges

As discussed largely in the scholarly community, the major challenge must be related to combining the two different methodologies. Based on philosophies positing that quantitative and qualitative research methods cannot be mixed due to fundamental differences between them resulting from diverse ontological, epistemological, axiological, and methodological assumptions, referred to as the “incompatibility thesis,” researchers advocating for this thesis have argued that MM research is not possible (cf. Howe, 1988; Johnson & Onwuegbuzie, 2004). Regarding the epistemology of MM research approaches, Teddlie and Tashakkori (2010) have argued that pragmatism constitutes the most suitable philosophical paradigm for MM research because it justifies the employment of multiple methods with different philosophical foundations as long as they produce satisfactory results to answer the research questions. Furthermore, adopting a pragmatic worldview as the philosophical basis, the research question is the decisive factor when designing a research methodology, not the paradigm from which the method derives (Teddlie & Tashakkori, 2010). In fact, to date, a number of methodologists have been advocating for a

departure from the “paradigm wars” (Johnson & Onwuegbuzie, 2004). Since the 1960s, a diverse range of authors in many disciplines, including education (e.g., Miles & Huberman, 1994; Strauss & Corbin, 1998), and researchers in applied linguistics (e.g., Dörnyei, 2007; Lazarton, 2005), have been actively suggesting MM research approaches.

What is more, there are some challenges that MM researchers need to face not only at the philosophical level but also at the practical level. That is, researchers have to carry out both quantitative and qualitative research together, obtain multiple data sets concurrently or sequentially, be cognizant about different methods and approaches, and understand how appropriately to mix them. Indeed, it is a daunting task for researchers to develop a solid understanding of two methodologies in order to use them appropriately in their research. Yet, although MM research possesses the strengths of both, it is also vital to consider that the users need to be familiar with the two methodologies, not just one or the other. For example, it is often the case with language assessment studies, many of which stem from a strong quantitative research orientation, that, although providing solid statistical analyses, sometimes, they have fallen short in presenting a rich qualitative picture of the phenomenon under study. Qualitative components tend to be used in a supplementary way to explain a phenomenon or test the effectiveness of an intervention. As seen in the example earlier, it is highly possible for qualitative research approaches to be the core of a MM study. Especially in classroom-based investigations that require rich description of a social interaction, MM researchers may wish to consider greater use of interpretative approaches in their studies.

In this sense, selection of an appropriate research design is another challenge for MM researchers. Over the last 15 years, many MM research designs have been conceptualized, and selecting the most optimal one from among these designs often is a challenging task for researchers. In their analysis of studies on the predictive validity of International English Language Testing System (IELTS), for instance, Lee and Greene (2007) have pointed out that a few early studies that employed both quantitative and qualitative research methods to collect data “failed to integrate findings from the different methods” (p. 369). Even when qualitative and quantitative research methods are used in a single study, without thoughtful design selection and integration of the data sets, one cannot achieve the tenets of MM research methodology and obtain optimal results. Thus, in order to answer their research questions, users need to know the basic principles of MM research designs to incorporate into their studies and how to integrate data sets from the two different paradigms appropriately. Indeed, there are methodologists who question the degree of data integration in some MM research studies as a quality standard for the methodology (e.g., Hashemi & Babaii, 2013); however, underpinned by successful exemplars, it is not doubtful that such an approach will continue to contribute to the field of language assessment and beyond.

Practical Challenges

There is also a practical and logistical issue that concerns researchers. Because of the approach involving several data collection stages, MM researchers need to develop multiple instruments, to gain access to participants for each stage, to collect and to analyze the data sets, and, finally, to integrate the entire data set in order to produce comprehensive results. From a CBLA research perspective, researchers have to be realistic and logistic about how much commitment the data collection phases would require, such as duration of data collection period in a classroom and numbers of students involved in different phases. Although it differs depending on a design, a MM research study tends to be heavy in terms of the data collection procedures. It might require a large number of people for the quantitative phase and an extensive duration of time for the qualitative phase. In this respect, the investigator might encounter more difficulties to recruit participants due to the complex data collection procedures. Although a research design should be exclusively determined by the research questions, it is also sensible to plan a MM research study within a feasible timeline and financial constraints.

Report writing and publishing are additional issues for MM researchers. Due to the nature of complex data collection procedures, analysis and integration, it is a major task to present a MM research study in an effective and audience-friendly manner (Sandelowski, 2003). Moreover, even though MM research has become more common in many disciplines including education, in language assessment, this approach does not seem to be as prevalent as do the other two paradigms as of yet. In an article on journal archive analysis, Chalhoub-Deville and Deville (2008) reported that the number of qualitative research articles published in the two major language assessment journals were increasing very slowly. Similarly, in the field of education, Stapleton and Collett (2010) in ELT, and Truscott et al. (2010) in general education observed a considerably small number of MM research study articles in their journals. Subsequently, Truscott et al. (2010) pointed out that researchers who hope to have their work funded or published might gravitate to quantitative research methods, which potentially contradicts our intention of making a contribution to science. Although this issue should not discourage researchers from using the methods that are most suitable for their academic inquiries, it is also true that knowledge dissemination is one of the main goals of conducting research. To create a more diverse research discourse from a publication perspective, MM research needs to be constructed in a scientifically valid and reliable manner in order to generate insightful knowledge that would not be possible through research conducted monolithically.

Conclusion: Potentials for Better Understanding Research Issues

Over the past five decades, the trends in language assessment have been shifting from measuring one-dimensional segmental linguistic skills to a more multi-dimensional, holistic approach (Bachman, 2007; Stoyhoff, 2012). The role of assessment in the language classroom has occupied an increasingly visible place in the language education literature as teachers, researchers, and test specialists examine how assessment can be woven into instruction to support language learning. Owing to the growing needs for research on this topic, research methodologies employed for CBLA research also have been more diverse and sophisticated. MM research approaches, which enable researchers to obtain knowledge that ranges from a general, macro picture to more detailed insights of specific cases, have become more visible in the field of CBLA for the past decade.

Despite the fact that language assessment researchers are increasingly finding that the relevance of research can be enhanced by MM research approaches, the field has not yet developed the academic discourse regarding the usefulness and applicability of this methodology for CBLA research. Given that the multi-faceted nature of the phenomena in classrooms requires an in-depth analysis of content-related variables, and that the measurement of the phenomena at the larger and wider level can be undertaken most effectively through quantitative research approaches, MM research approaches can help CBLA researchers obtain answers to the broader and more complete range of research questions. Although MM research approaches can provide a better understanding of the target issue in CBLA research contexts, it is essential to realize that there still are challenges with using this methodology.

The pragmatic principles of the new methodology have served as useful tools to investigate the complex nature of classroom assessment practices. Especially, MM research approaches seem highly useful for conducting research on the influential factors and efficacy of a new assessment practice/format in a language classroom at the same time. Through a holistic approach, this methodology has the potential to provide more credible findings than does a monolithic methodology when the data sets are appropriately integrated. Each methodology has its own strengths and weaknesses and MM research is no exception. Although there might be a variety of weaknesses in MM research approaches, when appropriate research design and context are chosen, MM research approaches are capable of helping researchers delve deeper into the research issue and contribute more thoroughly to their areas of scholarship.

References

- Anderson, N., Bachman, L. F., Cohen, A. D., & Perkins, K. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing*, 8, 41–66. doi:10.1177/026553229100800104
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1–42. doi:10.1191/026553200675041464
- Bachman, L. F. (2007, April). *Language assessment: Opportunities and challenges*. Paper presented at the annual meeting of the American Association of Applied Linguistics (AAAL), Costa Mesa, CA.
- Bailey, K. M., & Nunan, D. (Eds.). (1996). *Voices from the classroom: Qualitative research in language classrooms*. Cambridge, England: Cambridge University Press.
- Banerjee, J., & Luoma, S. (1997). Qualitative approaches to test validation. In C. Clapham & D. Banerjee, J., (Eds.), *Encyclopedia of language and education* (Vol. 7, pp. 275–287). Dordrecht, The Netherlands: Kluwer Academic Publishers. doi:10.1007/978-1-4020-4489-2_25
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning—putting it into practice*. Oxford, England: Open University Press.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80, 139–148. doi:10.1177/003172171009200119
- Boud, D. (1995). *Enhancing learning through self-assessment*. London, England: Kogan Page.
- Carless, D. (2006). *How assessment supports learning: Learning-oriented assessment in action*. Pokfulam, Hong Kong: Hong Kong University Press.
- Chalhoub-Deville, M., & Deville, C. (2008). Utilizing psychometric methods in assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 7, pp. 211–224). New York, NY: Springer Science + Business Media LLC.
- Cheng, L., & Fox, J. (2013). Review of doctoral research in language assessment in Canada (2006-2011). *Language teaching: Surveys and Studies*, 46, 518–544. doi:10.1017/S0261444813000244
- Clapham, C. (1996). *The development of IELTS: A study of the effect of background knowledge on reading comprehension* (Vol. 4). New York, NY: University of Cambridge Local Examinations Syndicate/Cambridge University Press.
- Colby-Kelly, C., & Turner, C. E. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. *Canadian Modern Language Review*, 64(1), 9–38.
- Collins, K. M., Onwuegbuzie, A. J., & Sutton, I. L. (2006). A model incorporating the rationale and purpose for conducting mixed methods research in special education and beyond. *Learning Disabilities: A Contemporary Journal*, 4(1), 67–100.

- Creswell, J. W., Fetters, M. D., & Ivankova, N. V. (2004). Designing a mixed methods study in primary care. *The Annals of Family Medicine*, 2(1), 7–12. doi:10.1370/afm.104
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage.
- Denzin, N. K. (2010). Moments, mixed methods, and paradigm dialogs. *Qualitative Inquiry*, 16, 419–427. doi:10.1177/1077800410364608
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford, England: Oxford University Press.
- Duff, P. (2006). Beyond generalizability: Contextualization, complexity and credibility in applied linguistics research. In M. Chalhoub-Deville, M., Chapelle, C., & P. Duff (Eds.), *Inference and generalizability in applied linguistics: Multiple research perspectives* (pp. 65–95). Amsterdam, Netherland: John Benjamins.
- Gass, S. M., & Mackey, A. (2007). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 175–199). Mahwah, NJ: Lawrence Erlbaum.
- Geertz, C. (1973). Thick description toward an interpretive theory of culture. In C. Geertz (Ed.), *The interpretation of cultures* (pp. 3–30). New York, NY: Basic Books.
- Ghoorchaei, B., Tavakoli, M., & Ansari, D. N. (2010). The impact of portfolio assessment on Iranian EFL students' essay writing: A process-oriented approach. *GEMA: Online Journal of Language Studies*, 10(2), 35–51.
- Gipps, C. (1995). What do we mean by equity in relation to assessment? *Assessment in Education*, 2, 271–281. doi:10.1080/0969595950020303
- Greene, J. C. (2007). *Mixed methods in social inquiry* (Vol. 9). New York, NJ: John Wiley & Sons.
- Guardado, M., & Shi, L. (2007). ESL students' experiences of online peer feedback. *Computers and Composition*, 24, 443–461.
- Gunning, P. (2011). *ESL strategy use and instruction at the elementary school level: A mixed methods investigation* (Unpublished doctoral dissertation). McGill University, Montreal, Quebec, Canada. Retrieved from http://digitool.Library.McGill.CA:80/R/-?func=dbin-jump-full&object_id=103480&silolibrary=GEN01
- Hashemi, M. R., & Babaii, E. (2013). Mixed methods research: Toward new research designs in applied linguistics. *The Modern Language Journal*, 97, 828–852. doi:10.1111/j.1540-4781.2013.12049.x

- Hamp-Lyons, L. (2000). Social, professional and individual responsibility in language testing. *System*, 28, 579–591. doi:10.1016/S0346-251X(00)00039-7
- Hamp-Lyons, L., & Lynch, B. K. (1998). Perspectives on validity: A historical analysis of language testing conference abstracts. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp. 253–276). Mahwah, NJ: Lawrence Erlbaum.
- Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational researcher*, 17(8), 10–16. doi:10.3102/0013189X017008010
- Ishihara, N. (2009). Teacher-based assessment for foreign language pragmatics. *TESOL Quarterly*, 43, 445–470. doi:10.1002/j.1545-7249.2009.tb00244.x
- Jamieson, J., & McDonough, K. (2012). *Quantitative and mixed methods: Overview*. In A. C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0988>
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, 30, 345–362. doi:10.1177/0265532213480334
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26. doi:10.3102/0013189X033007014
- Kikuchi, K., & Browne, C. (2009). English educational policy for high schools in Japan: Ideals vs. reality. *RELC Journal*, 40, 172–191. doi:10.1177/0033688209105865
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Lazaraton, A. (2005). Quantitative research methods. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 209–224). Mahwah, NJ: Lawrence Erlbaum.
- Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34–50. doi:10.1016/j.jslw.2012.11.003
- Lee, Y. J., & Greene, J. (2007). The predictive validity of an ESL placement test: A mixed methods approach. *Journal of Mixed Methods Research*, 1, 366–389. doi:10.1177/1558689807306148
- Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity*, 43(2), 265–275. doi:10.1007/s11135-007-9105-3
- Lumley, T., & Brown, A. (2005). Research methods for language testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 833–856). London, England: Lawrence Erlbaum.
- Lynch, B., & Shaw, P. (2005). Portfolios, power, and ethics. *Tesol Quarterly*, 39, 263–297. doi:10.2307/3588311
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18, 333–349. doi:10.1191/026553201682430076

- McNamara, T. (2003). Tearing us apart again: The paradigm wars and the search for validity. *EUROSLA Yearbook*, 3(1), 229–238. doi:10.1075/eurosla.3.13mcn
- Mezirow, J. (1996). Contemporary paradigms of learning. *Adult Education Quarterly*, 46, 158–172. doi:10.1177/074171369604600303
- Miles, M. B., & Huberman, M. (1994). *Qualitative data analysis: A sourcebook of new methods* (2nd ed.). Beverly Hills, CA: Sage.
- Mohan, B. A. (2012). Qualitative research methods in second language assessment. In L. V. Hedges, R. Coe, & M. Waring (Eds.), *Research methods and methodologies in education* (pp. 752–767). Thousand Oaks, CA: Sage.
- Moss, P. A. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher*, 25(1), 20–29. doi:10.3102/0013189X025001020
- Neumann, H. (2010). *What's in a grade? A mixed methods investigation of teacher assessment of grammatical ability in L2 academic writing* (Unpublished doctoral dissertation). McGill University, Montreal. Retrieved from http://digitool.Library.McGill.CA:8881/R/?func=dbin-jump-full&object_id=103454
- Piaget, J. (1975). *The child's conception of the world*. Totowa, NJ: Littlefield, Adams.
- Pollack, J. (2007). The changing paradigms of project management. *International Journal of Project Management*, 25, 266–274. doi:10.1016/j.ijproman.2006.08.002
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge, England: Cambridge University Press.
- Purpura, J. E. (2011). Quantitative research methods in assessment and testing. In C. Chapelle & E. Hinkel (Eds.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 731–751). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rocco, T., Bliss, L., Gallagher, S., Pérez, A., & Prado, P. (2003). Mixed methods taking the next step: Mixed methods research in organizational systems. *Information Technology, Learning, and Performance Journal*, 21(1), 19–29. doi:10.1.1.122.1050
- Rogoff, K. (1999). International institutions for reducing global financial instability. *Journal of Economic Perspective*, 13(4), 21–42. doi:10.1257/jep.13.4.21
- Sandelowski, M. (2003). Tables or tableaux? The challenges of writing and reading mixed methods studies. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 321–350). Thousand Oaks, CA: Sage.
- Scarino, A. (2013). Changing perspective on language learning: Towards multilinguality. *LEKSIKA*, 7(1), 12–19.
- Scriven, M. (1967). The methodology of evaluation: Perspectives on curriculum evaluation. In R. Tyler, R. Gagne, & M. Scriven (Eds.), *AERA monograph series – curriculum evaluation* (pp. 39–83). Chicago, IL: Rand McNally.
- Stapleton, P., & Collett, P. (2010). JALT Journal turns 30: A retrospective look at the first three decades. *JALT Journal*, 32(1), 75–90.

- Strauss, A. L., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.
- Stoynoff, S. (2012). Looking backward and forward at classroom-based language assessment. *ELT journal*, 66, 523–532. doi:10.1093/elt/ccs041
- Tan, M. (2011). Mathematics and science teachers' beliefs and practices regarding the teaching of language in content learning. *Language Teaching Research*, 15, 325–342. doi:10.1177/1362168811401153
- Tashakkori, A., & Teddlie, C. (Eds.). (2010). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.
- Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 3–50). Thousand Oaks, CA: Sage.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative techniques in the social and behavioral sciences*. Thousand Oaks, CA: Sage.
- Teddlie, C., & Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (pp. 1–44). Thousand Oaks, CA: Sage.
- Truscott, D. M., Swars, S., Smith, S., Thornton-Reid, F., Zhao, Y., Dooley, C., & Matthews, M. (2010). A cross-disciplinary examination of the prevalence of mixed methods in educational research: 1995–2005. *International Journal of Social Research Methodology*, 13, 317–328. doi:10.1080/13645570903097950
- Tsushima, R. (2011). *The mismatch between educational policy and classroom practice: EFL teachers' perspective on washback in Japan* (Master's thesis). Retrieved from http://digitool.Library.McGill.CA:80/R/-?func=dbin-jump-full&object_id=106524&silos_library=GEN01
- Turner, C. E. (2012). Classroom assessment. In G. Fulcher & F. Davidson (Eds.), *Routledge handbook of language testing* (pp. 61–74). New York, NY: Routledge.
- Turner, C. E. (2013). Mixed methods research. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1403–1417). Hoboken, NJ: Wiley-Blackwell.
- Turner, C. E., & Purpura, J. E. (2013, March). A proposed model of learning-oriented assessment (LOA). In J. E. Purpura, N. Saville, & C. E. Turner (Organizers), *ILTA at AAAL colloquium learning-oriented assessment (LOA) in classrooms: A place where SLA, interaction, and language assessment interface*. Symposium conducted at the annual conference of American Association for Applied Linguistics, Dallas, TX.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

- Wajda, E. (2011). New perspectives in language assessment: The interpretivist revolution. In M. Pawlak (Ed.), *Extending the boundaries of research on second language learning and teaching* (pp. 275–285). Heidelberg, Berlin, Germany: Springer.
- Weigle, S. C. (1994). Effects of training on raters of ESL compositions. *Language Testing*, *11*, 197–223. doi:10.1177/026553229401100206