

# **Section 1**

## **Language assessment and mixed methods research: Fundamental considerations and symbiosis**

Cambridge University Press  
978-1-316-50503-8 — Second Language Assessment and Mixed Methods Research  
Edited by Aleidine J. Moeller , John W. Creswell , Nick Saville  
Excerpt  
[More Information](#)

---

# 1 The confluence of language assessment and mixed methods

*Aleidine J Moeller*  
*University of Nebraska–Lincoln*

---

*Mixed methods generates results that are both smooth and jagged, full of relative certainties alongside possibilities and even surprises, offering some stories not yet told (Greene 2008:20).*

This chapter serves as an introduction to the convergence of language testing and assessment and mixed methods research. An examination of the benefits and value added of mixed methods research within the field of language testing and assessment is the primary focus.

Topics highlighted in this chapter include:

- The shift from a testing culture to an *assessment culture*
- An examination of concurrent parallel *paradigm shifts* within language teaching and learning, research approaches and testing and assessment
- The *history of language assessment* and its *connection to mixed methods*
- The fundamental principles of mixed methods research and its *application to language assessment*

## Introduction

The purpose of this volume is to create a deeper understanding of the role of mixed methods in language assessment; to provide fundamental information needed to conduct mixed methods research within the context of language assessment; to provide the reader with the essentials for conducting and publishing a rigorous mixed research study; and to expand the practice of mixed methods research into the field of language testing and assessment. As noted by Hashemi and Babaii (2013:828), mixed methods research has not fully gained the ‘kind of attention in applied linguistics research’ that it deserves, a sentiment extended by Turner (2014:1,413) when she opined that a greater

## Second Language Assessment and Mixed Methods Research

emphasis may need to be ‘placed on mixed methods research in language testing venues’.

This chapter further examines the benefits and value that can be added through mixed methods research within language testing and assessment. The field of language testing and assessment is informed by linguistics, applied linguistics, language acquisition and language teaching, as well as by the disciplines of teaching, measurement and evaluation (Shohamy 2010). Language testing basically consists of two fundamental components – ‘what’ is being tested and ‘how’, or which method is used for assessing the ‘what’ that is being tested (Shohamy 2010). During the last 50 years the language testing field has developed from discrete-point testing focused on lexical and structural issues to a more integrative approach that utilised authentic oral and written texts, and most recently to a communicative, task-based approach where language learners perform tasks within authentic, real-life contexts. As the definition of what it means to know a language has expanded and gained in complexity, language testers have sought to develop valid language assessment tools that represent varied and different societies in different contexts. Concerns have focused on political, social, educational and ethical dimensions, the meaning of language and ‘the possibilities for measuring this complex and dynamic variable’ (Shohamy 2010:xv).

This chapter begins with a discussion of the movement within the language field from a testing to an assessment culture and provides general background information and context regarding the concurrent parallel paradigm shifts from language teaching to learning, and from a psychometric to a more multiplistic research approach. Next, an examination of the impact of communicative language teaching and the move towards task-based instruction on language assessment is provided, together with a history of language assessment and its connection to mixed methods. This is followed by an introduction to the fundamental principles of mixed methods research and its roots in pragmatism, and a justification for its application to language assessment research. Examples of mixed methods designs that have added insight and a broader understanding to studies on the impact of language assessment are explored.

### **Moving from a testing to an assessment culture**

Spolsky (1977) identified three developmental stages of language testing in the 20th century: the pre-scientific, the psychometric-structuralist and the psycholinguistic-sociolinguistic. As large-scale testing gained popularity in the 20th century, statistical procedures became the primary evaluative procedures for test development and test evaluation (Kunnan 2008). Standardised norm-referenced tests were aimed at measuring a test taker’s language ability compared to other students and served to rank individuals for purposes of

### The confluence of language assessment and mixed methods

gate keeping. Such large-scale, norm-referenced tests produced a single score without benefit of meaningful feedback that identified a test taker's strengths and weaknesses in specific language domains. These tests provided little guidance for teachers to review curricular design and instruction in order to improve language learning. This approach to testing has been described as representative of a 'testing culture' (Wolf, Bixby, Glenn and Gardner 1991).

In the 1970s the sociolinguistic view of language testing as a means of interpersonal communication that is context related became the prevailing drive towards more integrative tests. As the concept of knowledge as a universal, fixed and measurable commodity gave way to the more process-oriented approach of 'seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there' (Assessment Reform Group 2002), learning became viewed as context bound, with teacher feedback playing a major role in supporting and promoting student learning (Black and William 1998, Inbar-Lourie 2010, James 2001, Shepard 2005). This view of assessment reflected Vygotsky's theory, which emphasised the cultural context of individual meaning making and Piaget's cognitive development theory, which regarded learning as 'an integral and inseparable aspect of social practice' (Lave and Wenger 1991:31). With this shift in paradigm from a testing culture to an assessment culture, assessment is now regarded as a value-embedded social activity (Filer 1995) with learning and assessment viewed as being inextricably linked. This new understanding of the interactive nature of learning and the role of assessment and assessors in the instructional-learning cycle (Black and William 1998) forced a reconsideration of the psychometric paradigm, which had dominated the testing cultures (Shepard 2000).

Cambridge English advanced an ecological model of Learning Oriented Assessment (Jones and Saville 2014) in which all levels of assessment, from the classroom to large-scale testing are 'brought into a complementary relationship and coordinated to provide maximum positive impact: providing evidence to support better learning, as well as better measurement and recognition of learning outcomes' (Jones and Saville 2014:6). While responsibilities for learning remain with the teachers and students in the classroom, there is a complementary relationship between large-scale and classroom assessment.

Investigating educational and social phenomena in context required insights from multiple disciplines using diverse research designs (LeCompte 2009). Social and behavioural sciences in educational research worked together as researchers in order to pursue crucial questions about 'how schools did or didn't work, why the best intended efforts of teachers often failed, and why students were not succeeding at desired levels' (LeCompte 2009:30). Gabriel Salomon (1991) argued that educational phenomena, like classrooms, are so complex they warrant the complementary use of

## Second Language Assessment and Mixed Methods Research

both ‘analytic’ and ‘systematic’ approaches and methods, perhaps most feasibly across studies. For example, speech acts and conversations are co-constructed by both interlocutors and reflect turn taking that cannot be fully captured in a discourse completion test (McNamara and Roever 2006). Given this complexity, a better understanding of the multi-faceted character of educational and other social phenomena can be obtained from the use of multiple approaches and ways of knowing. This multiple approach way of thinking yields a richer, deeper and better understanding of important facets of social and educational phenomena.

Enquiries into educational and social phenomena require exploration over time, in multiple settings and with a variety of informants. The complexity of such educational research cannot be overstated, as noted by Berliner, whose variation on the common phrase ‘It’s not rocket science’, acknowledged the complexities of researching teaching and learning: ‘Well, at least it’s not educational research’ (Berliner 2002).

### Concurrent parallel paradigm shifts

The field of language learning and teaching experienced a similar and significant paradigm shift, moving from a focus on teaching to a focus on learning, from achievement to proficiency, from textbooks to authentic texts and digital media and most recently, from methods to measurement (Lazaraton 2002); a shift that in many ways parallels a similar movement within language assessment as the field transitioned from a testing culture to an assessment culture (Inbar-Lourie 2010). This shift from a testing to an assessment culture meant moving from:

- assessing a specific skill to assessing language more holistically
- comparing student performance with another to comparing it with established criteria
- creating assessments independent of curriculum and instruction to those aligned with curriculum and instruction, the latter being a common practice in the UK where historically assessments were aligned with curricula and instruction
- making inferences based on single, restricted evidence to those based on multiple sources
- viewing students as objects of assessment to viewing them as active participants in the process
- holding a few responsible for assessment results to making all concerned with language learning accountable for assessment results.

Assessment culture, rooted in constructivist theories, underscores how knowledge is processed and developed and encompasses broader theoretical and practical frameworks for assessing knowledge (Inbar-Lourie 2010). The

## The confluence of language assessment and mixed methods

emphasis lies on the link between assessment and learning, reflecting a shift towards more inclusive, multiple forms of measurement that promote learning, rather than those that merely function as an audit of learning (Gipps 1994, Stiggins 2002).

Assessment culture embraces a broadening of the assessment construct by using assessment data from different sources and multiple informants and through multiple assessment tools that include various micro and macro sources; it emphasises multiple stakeholders playing an active role, and therefore becoming integral to the assessment process (Shepard 2000). Since assessment culture views assessment as a 'context-relevant activity grounded in learning, tests are sensitive to contextual variables, to the learners' cultural and linguistic backgrounds and to the knowledge they bring with them to the assessment encounter' (Inbar-Lourie 2010:296).

Consistent with the concept of an assessment culture, Purpura (2010) described the purpose of language assessment as collecting trustworthy, score-based and descriptive information about what students *know and are able to do with language*. Key questions to emerge, therefore, are: How can score-based inferences from classroom-based assessments be used to make decisions about student readiness to benefit from instruction, about student attainment and growth or about the kind of feedback to provide learners at different points in the learning process? How can teachers use this evidence to make decisions regarding next steps related to curricular content, instructional methods and classroom materials? What impact evidence do we have about the intended consequences of our classroom-based assessments for individual learners and teachers and how do these assessments serve to promote or inhibit further learning and more effective teaching? Such questions are not easily answered through a one-size-fits-all approach when assessing complicated phenomena and multi-dimensional outcomes in applied linguistics research. A mono-method approach may very well veil contributing factors to learner performance (Upshur and Turner 1999). Solano-Flores and Trumbul (2003) argue for new paradigms for assessing language learners in light of the complexities of language knowledge and the cultural issues embedded in language acquisition. Moving to a multiplicity of research methods, or mixed methods design, can better capture the inherent complexities involved in the assessment of language classroom teaching and learning.

## Communicative language teaching and language assessment

In order to fully understand the role of mixed methods in language assessment research, it is important to clarify how the shift towards communicative language teaching impacted assessment. With the move towards

## Second Language Assessment and Mixed Methods Research

teaching learners how to communicate in the real world, task-based language instruction, an approach that focuses on what learners can do with language, gained considerable momentum. Task-based assessment relies on meaningful, real-world, authentic language performance that should reflect the tasks and interactions that learners are likely to encounter in real-life situations. Van Gorp and Deygers (2014:591) note that task-based language assessment ‘provides a much-needed interface between theory and practice’, one predominantly dependent on implementation by the classroom teacher. Interest in task-based assessment has grown among language testers as it has proven valuable in raising awareness with stakeholders about language-learning processes. The numerous variables and complexity in assessing authentic task-based communication at the classroom level in addition to the challenges such as reliability, content validity and authenticity (Bachman and Palmer 2010, Norris 2009, Wigglesworth 2008) underscore that one research method cannot fully capture the complexity of language skills.

### **Mixed methods research and its application to language assessment**

Tashakkori and Creswell (2007) define mixed methods research as one in which ‘the investigator collects and analyzes data, integrates the findings and draws inferences using both qualitative and quantitative approaches or methods in a single study or program of inquiry’ (2007:4). Tashakkori and Teddlie (2008) posit that one of the strengths of mixed methods is that it brings together ‘information that can result in “*meta-inferences*” about the phenomenon under study that neither the quantitative nor qualitative perspectives could do alone’ (2008:101, our emphasis). They define *meta-inferences* as ‘an overall conclusion, explanation, or understanding developed through an integration of the inferences obtained from the qualitative and quantitative strands of a mixed methods study’ (2008:101). This underscores the innate connection between mixed methods research designs and the ‘*meta-inferences*’ they allow, and the ‘integrative evaluative judgement’ that test validity is (see Ziegler and Kang, Chapter 4, this volume).

A key feature of mixed methods research is its methodological pluralism, which provides a deeper and varied perspective versus a single research methodology. Johnson and Onwuegbuzie (2004) present mixed methods research as the ‘third research paradigm’, or a ‘third chair, with qualitative research sitting on the left side and quantitative research sitting on the right side’ (2004:15). They note: ‘the goal of mixed methods research is not to replace either of these approaches but rather to draw from the strengths and minimise the weakness of both in single research studies and across studies’ (2004:14–15).

When findings are corroborated across different approaches, there is



## The confluence of language assessment and mixed methods

greater confidence than through one approach alone; if the findings conflict, then the researcher has greater knowledge and can modify interpretations and conclusions accordingly. In many cases, the goal of mixing is not to seek corroboration, but rather to expand one's understanding (Onwuegbuzie and Leech 2004).

Khalifa and Docherty (see Chapter 11, this volume) provide an illustration of such an expansion of understanding in their test impact study, in which mixed methods extended the scope and validity of their research. The use of qualitative interviews and focus groups allowed for the inclusion of young learners as stakeholders whose voices could not be fully represented through a questionnaire. They also report how mixed methods enhanced the stakeholders' confidence in the findings as a result of the triangulation of information derived from multiple data sources. By utilising quantitative and qualitative techniques within the same enquiry question, mixed methods research can incorporate the strengths of both methodologies, providing a more defensible argument supported by richer information.

In order to investigate how language is learned and acquired in a variety of contexts, a research methodology commensurate with the research question is needed. The goal is to capture both external and internal assessments that measure knowledge, teaching, and learning in a natural setting and manner, and to analyse current practices and achievements that link assessment data in a socially interactive environment. Using a carefully chosen research methodology, a more complete picture of 'where the learners are in their learning, where they need to go and how best to get there' can be acquired (Assessment Reform Group 2002).

Mixed methods can be a challenging approach to research (Creswell and Plano Clark 2011). Guetterman and Salamoura (see Chapter 7, this volume) reviewed 10 mixed methods studies that represented a purposeful sampling of language assessment studies and examined the mixed methods features based on a checklist created by Creswell and Plano Clark (2011). They identify and describe nine methodological issues that emerged in these studies and how they negatively affected the validity of each study. They further discuss how to address each of these challenges through the addition of rigorous mixed methods and how to avoid these pitfalls when conducting language assessment studies.

## Pragmatism as a lens for mixed methods

Pragmatism is an emerging research paradigm focused on practice oriented understanding (Greene 2007) that emphasises what works (Creswell 2009). Pragmatism is well suited for mixed methods research as it can shed light on how research approaches can be mixed fruitfully (Hoshmand 2003) in ways that offer the best opportunities for answering important research

## Second Language Assessment and Mixed Methods Research

questions. Johnson and Onwuegbuzie (2004) argue for the use of pragmatism as the most effective mixed methods paradigm as it offers ‘an immediate and useful middle position philosophically and methodologically . . . and offers a method for selecting methodological mixes that can help researchers better answer many of their research questions’ (2004:17). The research question, rather than a pre-conceived worldview, is paramount and drives the choice of design. A key feature of mixed methods research is its methodological pluralism, which ‘frequently results in superior research (compared to mono-method research)’ (Johnson and Onwuegbuzie 2004:14). For a more detailed discussion of worldviews in mixed methods, see Ziegler and Kang, Chapter 4, this volume.

### **Mixed methods and the history of language assessment**

In the field of language teaching and learning, assessment has taken centre stage as high-stakes tests, as well as classroom assessments, have driven instruction and curriculum. A brief history of the evolution of language assessment will provide the requisite background for better understanding parallel developments in the use of various research methodologies as researchers grapple to find ways to increase student learning, improve instruction and seek the most appropriate and effective research methods for valid and reliable assessment of both.

The field of language testing in the 1970s moved towards a sociolinguistic view of language, one that is defined by purpose and context. Language testers moved away from an analytical approach to a more integrative testing approach (Davies 2014). In the 1980s, with the rise of the language proficiency movement, the scope of language testing was broadened ‘to bring it much more in line with other areas of applied linguistics’ (Skehan 1998:213). Skehan (1998) noted that this could provide language testing with the positive image it lacked, and ‘that tests would not always be done to people but with them’ (Skehan 1998:221). The 1980s also brought additional developments that contributed significantly to language testing. These included the application of item response theory (IRT), establishment of testing boards and agencies, increase in books on language testing, and the launch of the international journal *Language Testing*.

In the 1990s and into the following decade, the issue of washback emerged as a major concern. Alderson and Banerjee (2001) define washback as ‘the impact that tests have on teaching and learning. Such impact is usually seen as negative . . . however . . . a good test should or could have positive washback’ (Alderson and Banerjee 2001:214). Arguments about washback and impact fostered the International Language Testing Association (ILTA) to develop a code of ethics and a code of practice. During this same timeframe