

11. Assessing communicative competence

Luke Harding

0000-0001-9579-6571

Susy Macqueen

0000-0001-5404-6475

John Pill

0000-0002-8799-0267

Abstract

This chapter provides an overview and critique of communicative competence as it has been conceptualized and operationalized in language assessment. The chapter presents a historical overview of communicative competence in language assessment and a discussion of critical issues that demand attention from researchers and practitioners. We then examine communicative competence in relation to practical and theoretical developments in language assessment. A key concern for practitioners is that the act of assessing language forces them to prioritize certain communicative knowledges, behaviors, and patterns over others in a constrained sample of language use. A matter for theorists is that assessment practices themselves inhere a special kind of communicative competence and are worthy of investigation as anthropological and sociological phenomena. We provide examples of different methods that have been used to explore communicative competence in language assessment research. We conclude with recommendations for more inclusive directions in assessment design, a focus on the nature of communicative competence as it emerges in assessment contexts, and consideration of the affordances and challenges for assessment brought by technological advances.

The foundational premise of communicative competence in Hymes's (1972) paper "On communicative competence"—that the human capacity for communication comprises both language knowledge and the ability to activate it in alignment with usage patterns of relevant sociocultural contexts—has been an evolving concern for the field of language assessment since the communicative turn of the 1970s. The desire to understand the nature of communicative competence in assessment was driven by the need to describe the construct—the ability being assessed—as a precursor to operationalizing that construct through assessment tasks. Thus, language assessment researchers have not only drawn heavily on conceptualizations of communicative competence developed by Hymes (1972), Canale and Swain (1980), and Canale (1983), but have also themselves contributed to a wider understanding of communicative competence in the field of applied linguistics, most prominently in Bachman's (1990) and Bachman and Palmer's (1996, 2010) model of communicative language ability. However, understandings of communicative competence within the field of language assessment are hardly uniform, and the practical constraints of assessment, together with ideological orthodoxies embedded within many language assessment practices, have shaped the way in which communicative competence has been defined and understood over time and across contexts.

In the following sections, we set out the ways in which the field of language assessment has responded to the challenge of translating the theoretical notion of communicative competence into practice. Following a discussion of key conceptual developments and various translational models, we examine critical issues, discuss language assessment research methods, make recommendations for practice, and predict some future directions. An extensive literature providing critical perspectives on communicative competence in language assessment already exists (see, for example, Fulcher, 2000; Kramsch, 2006; Leung, 2005; McNamara, 1996; Spolsky, 1989). Within this chapter we connect with that tradition by using a critical lens to evaluate communicative competence within the assessment arena. Specifically, throughout the chapter, we argue, first, that the act of assessing language forces practitioners to prioritize certain communicative knowledges, behaviors, and patterns in a relatively constrained sample of language use. This act poses ongoing challenges not just for the practitioners themselves, but also for anyone using assessments in classrooms, institutions, and policies. Second, we propose that assessment practices themselves inhere a special kind of communicative competence and are worthy of investigation as anthropological and sociological phenomena.

Historical context

Communicative competence was primarily a theoretical concept arising within linguistic anthropology and not a practical framework for language learning, pedagogy, or assessment. We do not intend to outline foundational papers in depth (e.g., Hymes, 1972; Saville-Troike, 1982), as they have been comprehensively discussed in other chapters in this volume (see Chapter 1). However, it is important to establish that in these early discussions, critiques of testing and assessment played a prominent role. Hymes's problematizing of Chomsky's "ideal speaker-listener" in the "completely homogenous speech community" (Chomsky, 1965, p. 3), for example, arose out of his concern for what such a definition of linguistic competence meant for "real children" (Hymes, 1972, p. 270) whose competence would be considered, unjustly, at odds with the competence valued by institutions. Hymes observed that, "given subcultural differences in the patterns and purposes of language use, children of the lower status may actually excel in aspects of communicative competence not observed or measured in the tests summarized" (p. 274). The limitations of the *tests*, the paucity in scope of the ability measured, formed a central plank in Hymes's critique. Similar concerns about the misrepresentation of communicative competence in tests have also been discussed by Saville-Troike (1982) and Milroy and Milroy (2012).

It is therefore worth noting, from the outset, that language assessment has always had an uneasy relationship with communicative competence. Despite the strong influence of communicative competence on current models and theories in the field, language assessment remains a crucial site at which more expansive theories often collide with the reductive and restrictive concerns of standardization and consistency of measurement. In this section, we chart how communicative competence was translated into the field of language assessment and how recent research has challenged existing orthodoxies.

Translating communicative competence into assessment practice

Over the last 50 years, applied linguists, including language testing specialists, have developed translational tools for the purpose of converting theories of communicative competence into assessment practice. These translational tools fall into two broad,

interrelated categories: (1) theoretical constructs in the form of models and frameworks, and (2) fields of practice which comprise approaches to assessment and their associated methods. *Theoretical constructs* have become increasingly elaborate as understandings about various aspects of communication and language ability have developed (Macqueen, 2022). Key among these are the models of “communicative competence” developed by Widdowson (1978), Canale and Swain (1980), Canale (1983) and Celce-Murcia (2008), and a model of “(communicative) language ability” developed by Bachman (1990) and Bachman and Palmer (1996, 2010). *Fields of practice*, which either arose from the theoretical modeling or became strongly associated with it, include “communicative language testing” (Morrow, 1979, 2012), “performance testing” (McNamara, 1996), “specific purposes testing” (Douglas, 2000), and “task-based language assessment” (Brindley, 1994; Mislavy et al., 2002). These approaches have tended to value tasks and scoring methods (criteria, rating scales) designed to be relevant to the target domain of language use.

One of the most developed theoretical models is the “theoretical framework of communicative language ability” proposed by Bachman and Palmer (1996, p. 84). In its most recent form, set out in Bachman and Palmer (2010), the construct “language ability” is defined as “a capacity that enables language users to create and interpret discourse” (p. 34). Language ability is broken into two central attributes or sub-constructs: *language knowledge* (stored “pragmatic” and “organizational” language information) and *strategic competence* (a set of metacognitive strategies used when mobilizing language in situ). These two central components can be traced to Canale and Swain’s (1980) original framework, which separated language-related aspects from the ability to, for example, strategically manage a communication breakdown. Celce-Murcia (2008), by contrast, made strategic competence an all-encompassing set of behaviors that facilitates both communication (e.g., negotiating meaning) and learning (e.g., memory strategies). Canale and Swain also differentiated the structural aspects of language, such as morphosyntax, lexis, and phonology, from their sociolinguistic conventions, a distinction Bachman and Palmer maintain by dividing language knowledge into *organizational* knowledge and *pragmatic* knowledge.

In addition to these key ingredients, Bachman and Palmer (2010) propose that individuals draw upon *personal attributes* (e.g., age, personality, educational experience) and *topical knowledge* (i.e., information base). Actual performance is executed through the use of *cognitive strategies* (e.g., making associations, applying rules), and it is filtered through *affective schemata* (feelings associated with topics). When individuals use language, the attributes interact with one another, intra-individually (e.g., topic knowledge and language knowledge), inter-individually (e.g., test taker and examiner), and/or with “characteristics of the situation” (e.g., texts, tasks, technology, p. 34). Context is therefore a crucial element.

Despite translational efforts such as Bachman and Palmer’s communicative language ability model and well-developed traditions of scholarship to guide fields of practice, test developers have tended to mobilize communicative competence in test infrastructures (tasks, items, rating scales, scoring, etc.) through relying on even more straightforward and practical instruments. One widely used framework is the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001, 2020), which comprises lists of scales describing what language learners “can do” across different levels, contexts, and modalities. Although more useable, such instruments tend to simplify and essentialize theoretical tenets, potentially sacrificing the richness and complexity of the construct (Harding, 2014). Yet instruments like the CEFR—and comparable frameworks such as the American Council on the Teaching of Foreign Languages (ACTFL) scales and China’s Standards of English

Language Ability (CSE)—appeal because they make the process of test development more manageable and test scores interpretable for various stakeholders. As such frameworks gain recognition among educators and policy makers, they become *de facto* constructs themselves, thus creating conditions where language test providers have to demonstrate their tests' alignment with the framework to gain recognition. Such frameworks are intended to be malleable (see Deygers et al., 2018), and updates to the CEFR in particular have embraced wider concerns such as plurilingualism and mediation (Council of Europe, 2020). However, there is a limit to the extent to which any framework of this kind can be adapted for specific contexts of use (see Brunfaut & Harding, 2020). In this way, the process of translating a theoretical model can (and often does) become a process of simplification and standardization (McNamara, 2011).

The problems of performance and context

Even within more simplified construct definitions and operationalizations of communicative competence, there has been a clear shift in the field of language assessment towards explicitly acknowledging context in test tasks, towards more authentic tasks, and towards valuing communicative effectiveness in addition to—or as a superordinate criterion for—linguistic competence (i.e., knowledge of phonology, lexis, syntax, etc.). This “mainstream” approach (Harding, 2014) to communicative language assessment is broadly in alignment with the various theoretical models which, to different degrees, include (1) a store of language knowledge and (2) an ability to mobilize it in performance (see McNamara, 1996, for further analysis of key models). Developments in theorizing interactional competence have helped to form a clearer view of the dynamics of interaction in the test construct (e.g., Galaczi & Taylor, 2018; Ross, 2018), and language assessment researchers have become increasingly interested in more sophisticated, meaning-based approaches (e.g., Purpura, 2017). Yet there remain complex, perhaps intractable, problems in the mainstream communicative approach with respect to how we view “performance.” Indeed, performance has been a topic of sustained discussion in the field for over 25 years.

An individual's performance, or, to use Hymes's term, the *actual language use*, that emerges in the moment of testing and in the experience of the test taker is the *operationalized construct* (Knoch & Macqueen, 2020; Macqueen, 2022). In language test design, the test taker's performance is intended to reflect the types of communicative abilities deemed important and relevant to a target domain of language use. Yet, the test taker's performance has been manufactured through societal processes, test methods, and understandings and expectations about what to do when being assessed, among other things (Chalhoub-Deville, 2003; McNamara, 2007). This interrelationship with the context of the test itself makes the performance a particular sort of artifact that is distinct from the test taker's communication in non-assessment circumstances. Thus, we encounter the perennial problem for language assessment of generalization from particular, manufactured samples to future *actual language use*. We return to this problem in the “Critical issues” section.

Communicative performance is also known to be affected by the specific dynamics of the test taker and the assessment context. Among the many potential interacting factors are the test instruments (e.g., rating scales, task types); the particular test task, topic, and version (e.g., informal conversation about holidays, formal essay on government surveillance, pair discussion on food preferences); the human interlocutor/rater (e.g., personality, gender, interpretation of and experience with the rating scale, language ideologies); the machine interlocutor/rater (e.g., degree of interactivity, scoring algorithm and its input data, training

mechanisms, extent of human oversight); the test taker's current state (e.g., tired, stressed, confident); and multiple other factors (e.g., the position of the computer, the size of the room, the audibility or clarity of the instructions, the actions of the test administrators; for overviews see Fulcher, 2003; Knoch; 2022; Nakatsuhara et al., 2022). A significant challenge in performance assessment is the paradox of viewing communicative language ability as an *individual* ability but one that we can only observe in a *social* context. This has led to a deep theoretical question—"whose performance?" (McNamara, 1997)—the answer to which has practical implications for assessment design and scoring procedures.

Critical issues

Throughout the historical trajectory described in the previous section, the field of language assessment has grappled with a range of fundamental challenges related to conceptualizing and operationalizing communicative competence. Many of these challenges remain unresolved. In this section, we outline three critical issues: (1) Understanding test-taking competence as a type of communicative competence; (2) Negotiating the scope of communicative competence; and (3) Dismantling boundaries and resisting reified norms.

Test-taking competence as a type of communicative competence

One of the main purported benefits of the shift towards communicative approaches in language assessment is the central role of authenticity. For example, in communicative tasks for receptive skills assessment, test takers may be asked to read or listen to "authentic" texts drawn from real-world materials or to respond to texts in ways that simulate language activities in the target domain. In task prompts for productive skills, the intended reader/listener and the purpose for the performance will be specified—as these elements would be known in real-world settings—and a time limit set to approximate real-life demands of task completion. Criteria used to assess test performances will commonly cover features in addition to phonological, lexical, morphological, and syntactic knowledge, including, for example, task achievement, appropriateness, turn-taking, and promptness of response (in oral interaction) or discourse management (in writing). The underlying motive is that, in striving for authenticity in the test setting, a more accurate simulation of real-world communicative competence can be observed (see Norris, 2016). However, as we have noted in the previous section, taking part in a language exam is a kind of "staged" performance (Rydell, 2015, p. 535) which requires its own form of communicative competence. Tacit expectations about "good" test performance might be more or less well understood by different test-taking populations (for example, test takers with low print literacy in their first language might start from a position of relative disadvantage, see Deygers et al., 2021). Indeed, "appropriate" test-taking behavior may require overt or covert socialization, and this creates a number of problems for interpreting test performance.

Viewed from this multidimensional perspective, the connection between communicative competence and authenticity becomes more complex. A richly contextualized speaking role-play between a human interlocutor and a test taker is considered more "communicative" in the theoretical sense than a multiple-choice item on a reading text, even though in practice both tasks require a sophisticated knowledge of the context of use for successful performance. The multiple-choice item, for all its decontextualized appearance, requires familiarity with the nature of tests and their peculiar conditions, an understanding of the text type, the nature of the writer, the intended audience of the text, the topic content, the format of questions, the interplay between possible answer options, the likely knowledge/skill being

elicited, and the valued language variety, among other task features. From the point of view of anthropological linguistics, these are all relevant to communicative competence, but they may be taken for granted in interpreting performance on a multiple-choice test, as many individuals are socialized into these natural-seeming practices at a young age through schooling.

Relatedly, although the term “performance testing” is applied specifically to test designs that prioritize more direct and authentic sampling (e.g., a simulation of a pilot speaking to an air traffic controller in a test of Aviation English), any test (other than covert observation) is actually a performance, because the test taker understands that they are producing a sample of language for a particular purpose (i.e., to perform in the role of test taker and do test-taker activities). In tasks that also specify or imply an audience (e.g., a role-play where test takers might be required to display authority or empathy, or a letter writing task where their language is for a particular recipient), the test taker has a dual audience to consider: *the task audience* and *the judge*, which may be a computer algorithm or a human. In assessment contexts, understanding the nature of the simulation context and, in some simulation contexts, performing for two audiences, form part of the communicative competence peculiar to successful test taking.

The issue of tacit expectations is perhaps most complex for tests administered internationally. Developers creating tests for a particular domain (e.g., readiness for university study in English in the UK) may include features in their test tasks that they assume are universal while, in fact, these features operate differently in other contexts. For example, the value given to sharing personal experience in seminar discussion may not be the same in all academic cultures; expectations of who can initiate a conversation, change its topic, or disagree with a proposition may vary (Toomaneejinda, 2018). It is reasonable that test takers should demonstrate the knowledge and skills (that is, the communicative competence) to perform appropriately in the new context. However, test takers’ possibly substantial efforts to make the required cultural shift risk going unrecognized and unrewarded because this aspect of competence has been taken for granted in the test design. The particular socialization required for communicative competence in language test settings remains a sorely under-researched topic, although the study of washback and test impact—specifically, the effect of a test on the preparation of intending test takers and their teachers—offers a glimpse of this (e.g., Macqueen et al., 2019). More than simply the mobilization of test-taking strategies, test preparation socializes test takers into ways of being with test-induced values, for example, projecting an ideologically favorable stance on being a “good immigrant” (Rydell, 2015, p. 543). At a societal level, individuals are socialized into evaluating themselves through internalizing assessment criteria, propelled by the understanding that tests are a gateway for opportunity and self-improvement (Broadfoot, 1996).

Negotiating the scope of communicative competence

A second critical issue for language assessment concerns the need to reconcile different perspectives on the scope of communicative competence—where to draw boundaries and whose views should take precedence. Examples from the field of testing language for specific purposes (LSP) illustrate how different views of the scope of communicative competence may come into conflict. In 2017, a news article reported the complaint of a nurse trained in the Philippines that the writing task in the test of academic English required as part of her professional registration in the UK did not appear to match the communicative demands of the healthcare workplace. In “Do I have to understand jam-making to be a nurse?” (Pym,

2017), the headline refers to the nurse's dismay at being asked to analyze a diagram about the process of jam-making and describe this in an International English Language Testing System (IELTS) writing task. While English language teachers will recognize the skill of describing a process in writing as being useful to nurses (as well as to students and professionals in other disciplines), the task had little relevance for the test taker. For her, the necessary communicative competence was more specific. Given this reaction, researchers may argue for the suitability of a language test explicitly designed for healthcare professionals, such as the Occupational English Test (OET; occupationalenglishtest.org), originally developed for use in Australia (see McNamara, 1996).

Limits on the scope of the construct recognized in a LSP test such as the OET, however, are also apparent in a second example. Under Australian federal law, English language proficiency and professional competence must be assessed separately for the professional registration of health professionals trained outside the country (McNamara, 1996, p. 40). As such, the OET is designed to include features relevant to healthcare settings (e.g., topics, texts, interactional modes, criteria; thus targeting communicative competence in clinical situations), but not to test clinical competence itself. Drawing the boundary is difficult, though, as domain-specific communicative competence can become challenging to separate from a wider definition of clinical competence. Stakeholders may view aspects of communicative competence as belonging to their domain to teach and assess. To illustrate, a representative of the registration board for medical practitioners in Australia, when explaining how the board viewed its use of the OET, stated, "We don't think it's testing clinical communication skills.... If testing is congruent with practice[,] that's terrific, but we shouldn't be relying on that as the method for saying...these people will be good clinical communicators and culturally competent" (Macqueen et al., 2021, p. 8). In the representative's view, assessing communicative competence for a clinical setting (e.g., the doctor-patient consultation) is more than can be expected of a language test.

This example indicates how attempts by test developers and researchers to broaden the scope of the construct of a language test could be resisted by subject-matter experts and other stakeholders wanting to limit it to more traditional "language" features and unwilling to cede control of aspects of communicative competence that they view as professional (rather than language) skills. Other research studies attempting to expand the construct of LSP tests to encompass professional skills performed through language have shown that seeking a border between language and content can be contentious (e.g., for trainee teachers using a foreign language to teach mathematics and science, see Elder, 2001; for medical practitioners working in a language different from that of their training, see O'Hagan et al., 2016). Communicative competence makes greatest sense in terms of specific tasks and contexts, but, from a language assessment perspective, it is here that it becomes most difficult to establish boundaries for the construct and to generalize performance to other settings—that is, to predict the quality of performance in the domain from restricted information in test performance (see Chalhoub-Deville, 2003). A further challenge for assessment concerns the personnel to apply the test criteria specified (Elder et al., 2017). Language-oriented examiners may make inadequate proxies for participants in the domain, ignoring features that are valued and rating unfamiliar aspects of performance inconsistently, whereas subject-matter experts may give less attention to conventional linguistic criteria. Then, when test performances are rated by a subject-matter expert and a language expert together, further questions arise about whether the experts should use separate rating schemes and about whether a single overall score should be determined or component scores reported presenting the two perspectives. Exploring the scope and profile of communicative competence from

different viewpoints—including those of language experts, subject-matter experts, and test takers—is essential to establish which aspects matter to such groups and to consider whose values should be reflected in test design.

Dismantling boundaries; resisting reified norms

While communicative competence by default challenges the notion of idealized codes, models for teaching, learning, and assessment have tended to keep single-language boundedness as a fundamental property (e.g., communicative competence in *English*, where “English” is a single, standardized variety). Recent shifts in the wider field of (applied) linguistics have challenged this view in two ways. First, the phenomenon of “a language” is no longer seen as a bounded and inevitable arrangement of rules and patterns but rather a sociopolitical construct that does not match the everyday reality of multilingual speakers (e.g., Makoni & Pennycook, 2007). Second, and directly related to this, beyond the field of language testing, developments in the nature of communicative competence have drawn attention to the primacy of context in the mobilization of linguistic repertoires (e.g., Blommaert et al., 2005; Kramsch & Whiteside, 2008) and to the problem of using monolingual, native-speaker competence as a yardstick for users of multiple languages whose knowledge comprises “dynamic constellations of resources” that emerge in interaction from internal and social processes (Hall et al., 2006, p. 229; see also Chapter 3, this volume).

The challenges to communicative language assessment raised by these critiques are stark. The field has only begun to address the inherent instability and dynamic nature of lingua franca communicative environments (Canagarajah, 2006; Harding & McNamara, 2018) and to explore constructs of multilingualism in language assessment (Schissel, Leung, & Chalhoub-Deville, 2019; Shohamy, 2011). There has been some progress with respect to removing the native speaker as a benchmark in performance descriptors (e.g., Council of Europe, 2020), specific innovations in local testing contexts (e.g., Motteram, 2020), and research agendas seeking to provide an empirical basis for change (e.g., Ockey & Wagner, 2018). Yet large-scale communicative language assessment often remains mired, operationally, in standard language ideology; linguistic patterns that are valued in tests tend to reflect high-status, official or standard varieties that may not represent the dynamic reality of the target language use domain. As a case in point, tests of English language proficiency, for instance, are now used widely for admissions purposes for English as a Medium of Instruction (EMI) institutions in non-English-dominant contexts. Such tests will often be built on the norms of standardized varieties (e.g., featuring speakers of North American, British, or Australian English), but the test results may be generalized to situations where varieties of English are more heterogeneous or where translanguaging is a frequent and natural feature of communication (see Iliovits et al., 2022, for examples of language use at the American University of Beirut). Not only do such tests not elicit performance samples that match the target language use domain, they implicitly reinforce inappropriate monolingual standards in heteroglossic situations (Schissel, Leung, & Chalhoub-Deville, 2019).

Newer conceptualizations of communicative competence might address these dilemmas. For example, Leung (2005) argues that the theoretical notion of communicative competence should be recast as an “ethnographic orientation that is capable of making connections with emergent sensibilities in diverse contexts of English language learning, teaching and use” (p. 121). Concepts such as translanguaging (Canagarajah, 2018) have broadened the scope of communicative competence to include “more expansive spatial repertoires that transcend text/context distinctions and transgress social boundaries...to also treat meaning making

ability as distributed, accommodating the role of social networks, things, and bodies, beyond mind and grammar, requiring strategic emplacement” (p. 52). Such ideas create substantial challenges for operationalizing the assessment of communicative competence in ways that are both practical and fair to test takers from diverse backgrounds. And yet, getting the construct of assessments right is not only a matter of theoretical coherence, it is also a matter of social justice. Taking a decolonial perspective, García and colleagues (2021) focus on the “vast linguistic complexity and heterogeneity of people and language” (p. 3) and argue that a consequence of assessing a racialized bilingual child’s ability via the norms of single, named languages is that the child is frequently deemed linguistically deficit (see also Shohamy, 2001b). The concept of communicative competence originated in observations, in classrooms and elsewhere, that an idealized linguistic competence was simply not adequate to provide a full understanding of the human capacity for language or, with real-world consequences, to acknowledge the equal status of children in institutions with which their divergent linguistic repertoires did not align. Achieving social justice while balancing other fairness considerations (Kunnan, 2018) within the practical constraints on any assessment process is likely to remain one of the most complex challenges for language assessment into the future.

Research methods in assessment

As stated earlier, language assessment specialists have frequently drawn on theoretical models of communicative competence to guide their thinking in defining constructs of interest. These theoretical accounts have typically been supplemented, or in some cases enhanced, by empirical approaches that have sought to determine the nature and scope of communicative competence in specific contexts and to identify authentic tasks that interactants frequently engage with in the target language use domain. This empirical approach is known as the *domain analysis*. Domain analysis refers to the “forms, meanings, and use, assemblies of knowledge, skills, and competencies that language learners typically engage with in target language use situations, and features of tasks that invoke them” (Yin & Mislevy, 2022, p. 291). The domain analysis provides an important starting point for constructing an assessment that taps into elements of communicative competence valued in real-world settings. The domain also provides an important point of comparison in carrying out validation research. Ideally, just as test tasks should represent key aspects of the domain, test performance on those tasks should indicate future performance in that domain (Chapelle et al., 2008; McNamara, 2000). Thus, comparing the nature of the domain with characteristics of performance elicited from test tasks is essential if claims are to be made that a test measures communicative competence as defined in a given setting.

Various research methods have been used to analyze language use domains, characteristics of test performance, and the relationship between these. Methodological variety is warranted because the focus of analysis may be on the nature of language use (ranging from lexicogrammatical features to broader discourse categories such as register and genre), interactional patterns, tasks, texts, modalities, and other key features of communication within a given domain. In Table 11.1 we provide an overview of four common research methods currently used in assessment research with a focus on exploring aspects of communicative competence, providing a brief explanation of each method and what it might be most useful for, and providing examples of prototypical studies for follow-up reading.

Method	Description	Prototypical studies^a
Discourse analysis / Conversation analysis	Analyzing discourse—whether the discourse characteristics of real-world communicative settings or of test-taker performance—is a widely used method in assessment research. This is particularly the case in research focusing on pragmatics or interactional competence, where tools from conversation analysis have been applied to understand a range of phenomena in test-taker talk.	Brown (2003), Roever & Kasper (2018), Youn (2020) See also Chapter 8, this volume
Corpus linguistics	Corpus-based approaches are becoming increasingly common in language assessment research. For more communicatively oriented assessment research, corpora provide opportunities for insight into real-world communication (through existing or specially collected corpora) and into test-taker discourse (through corpora constructed with speaking or writing performances drawn from tests).	Gablasova et al. (2017), He & Dai (2006), Staples et al. (2017) See also Chapter 7, this volume
Ethnographic approaches	Ethnographic approaches may involve participant or non-participant observation of language use in a given domain or may attempt to understand the domain through close consultation with domain experts. As Douglas has suggested, ethnographic approaches constitute “an approach to describing and understanding a target language use situation from the perspective of language users in that situation” (2000, p. 93). Ethnographic approaches may also be useful for exploring the test event itself (seen next section).	Elder et al. (2017), Iliovits et al. (2022), Leung & Lewkowicz (2013) See also Chapter 5, this volume
Verbal report methods	The verbal report method includes a range of approaches such as concurrent and retrospective think-aloud protocols, as well as stimulated recall. These methods are useful for understanding the response processes of test takers on particular tasks, enabling comparison with processes required in real-world communicative settings. The same methods are also commonly employed to explore what aspects of communicative competence raters notice and value in performance assessment.	May (2011), Plakans (2009), Schissel, López-Gopar, et al. (2019)

^aIn several cases, multiple methods are used in the same study.

Table 11.1. Four common research methods for exploring communicative competence in language assessment

Recommendations for practice

At this point in the chapter, a practitioner might wonder how the more intractable issues can be addressed. Here we make two recommendations for practice emerging from the foregoing discussion: (1) refocusing assessment design on repertoires of communicative competence, and (2) shifting research priorities towards understanding the nature of communicative competence within test situations themselves.

On the first point, language assessment—whether for large-scale, international purposes or at the classroom level—can only proceed with a clear purpose and construct definition. The fact that theories of communicative competence have become more complex makes the task of operationalizing them more challenging, but it remains possible to address this challenge with the tools and concepts that are currently available. As Harding and McNamara (2018) have argued with respect to designing assessments of lingua franca competence, “language testing research has a history of identifying and solving problems in communicative language assessment on which it can draw” (p. 575). The primary task is to articulate a revised construct of communicative competence that can function as a foundation. The second task is to translate that construct into a set of assessment methods and procedures that captures the construct effectively. Given the increasing focus on repertoires of multi/translingual performance, it is likely that meeting this challenge will result in very different kinds of assessment tasks, for example, highly interactive, focused on accommodation and negotiation of meaning, where there are porous boundaries between named languages and dialects, and where communicative appropriateness or effectiveness is determined in context by the participants themselves. It is clear that a shift towards capturing these new conceptualizations of communicative competence would necessarily destabilize the current monolithic approach to large-scale, international testing, and would require instead a shift towards locally developed assessments, designed in consultation with learners and other stakeholders, and with democratic principles of test-taker agency at the fore (Shohamy, 2001a). In this regard, meeting the challenges of newer theories of communicative competence requires both technical, design-based change and systemic change in the sociopolitical and economic orthodoxies of language testing.

With respect to test taking as a kind of communicative competence, the field would benefit from recognizing this fundamental problem more openly, acknowledging that the test situation is a communicative event in its own right with its own properties. We must therefore recognize the additional *layer* of communicative competence that is required to take part in the test itself. In some cases, this “test-taking communicative competence” will need to be made more transparent, because what is taken for granted is often the source of inequalities (such as the child who has never had experience of transferring answers to an answer sheet, or who does not feel it is appropriate to question an adult in an oral proficiency exam). Minoritized learners whose existing communicative competence is currently not valued in test constructs and who have not been socialized into the particular communicative competence of “taking the test” are doubly penalized.

Addressing this issue would require a greater shift towards research on the test as a specific site of communication. There has already been a long tradition of such research with a focus on oral proficiency interviews (OPIs), with early work identifying such exams as a type of “institutional discourse” that is distinct from regular conversation (e.g., Young & He, 1998; see McNamara & Roever, 2006, for an overview of this research). One recent study has broadened these concerns to consider the dimensions of ideology and power that influence discourse in speaking test environments (Rydell, 2015). However, there has been less research on other aspects of test taking as having distinct communicative identities and relatedly little attention on the communicative competence required to engage with a wider range of test tasks. Beyond language assessment, work has been conducted in this vein by Maddox et al. (2015) from an anthropological perspective on the UNESCO Literacy Assessment and Monitoring Programme (LAMP) tests, with a focus on understanding the taking of a test “ethnographically as a distinctive social occasion” (p. 296). An ethnographic

approach is valuable, as the authors state, because it, “provides qualitative insights into how the test and test items are received and understood by the tested population. This is particularly important because ‘realistic’ test items carry tacit cultural knowledge and assumptions that may not be shared or understood by the tested population (e.g., scope for ‘cultural misfit’)” (pp. 296-297). We look forward to more research focusing on different aspects of language assessment to identify components of the layer of communicative competence that is required to engage successfully with a range of test-taking situations.

Future directions

The most urgent challenge on the horizon for language assessment, arguably, is dealing with the disruptive influence of digital technology on all aspects of language assessment practice and on the communicative practices that language assessments seek to measure (see Chapter 10, this volume). New language constructs require a reconceptualization of communicative competence to capture hitherto under-researched forms of engagement with digital tools. Such communication often blurs distinctions traditionally made between spoken and written modes to create its own discourses and genres. For example, text messages are typically informal and may include abbreviations, images, and emojis (without meaning for outsiders). Chains of messages are co-constructed by multiple participants, as if in spoken conversation. Readers post responses and discuss online articles creating nets of interacting opinions. Participants in a videoconference contribute to spoken discussion while commenting in parallel using the written chat function (and perhaps concurrently sending messages in private conversations using other software). Fleeting comments become permanent and retrievable unless designed to expire after receipt. Writers contribute to and edit the same text simultaneously online. In addition to reconceptualizing current constructs, the promise of technology for language assessment opens up new vistas for narrowing the gap between test settings and real-world communication. Virtual reality, for example, would allow for a fully immersive experience such that the communication required in a simulated assessment situation would match more closely the communicative competence of real-world domains of language use. However, it remains to be seen just how such settings—and, once again, the knowledge that one is being assessed—would influence the kind of performance elicited.

The more immediate challenge, however, is the threat that technology in test design and administration may narrow or dilute the more expansive definitions of communicative language ability that have been developed over the past 40 years. As we previously noted, theoretical models of communicative competence have advanced to a point where learning, teaching, and assessment are compelled to focus on linguistic repertoires in contexts of heteroglossia. Yet many online language assessment systems, while providing cheaper, more flexible, and more accessible experiences for test takers, are nevertheless less capable of tapping into these more sophisticated competences (see also Harding & Fulcher, 2022), particularly due to the absence of a human interlocutor (Roever & Ikeda, 2022) and the current limitations of automated scoring systems to process unpredictable spoken and written performances (Isaacs, 2018). The present situation represents an interesting tension for language assessment, and it is in this space that future innovations in communicative competence for assessment are likely to be located.

Sample test materials and discussion questions

The following role-play task might be used in an English-language proficiency test taken as part of the requirements for professional registration of nurses in an English-dominant context. Read through the task and consider the discussion questions below in light of the themes covered in this chapter.

Example task A

Test taker (NURSE)

You are working on a hospital ward. The 65-year-old patient, who is very overweight, had a serious fall and broke their leg. The patient is due to go home today. Your task is to help them consider eating a healthier diet as they continue to recover at home.

- Find out about the patient's eating habits.
- Encourage the patient to make healthy food choices.
- Offer suggestions that achieve your goals and suit the patient's situation.

Interlocutor (PATIENT)

You are a 65-year-old retired teacher, in (the) hospital after having a fall in which you broke your leg. You are overweight but feel fine for your age. You are going home today. The nurse is to give you advice about eating a healthier diet.

You have never liked vegetables or salad. You live alone and don't enjoy cooking. Shopping and preparing food are a waste of time. Fast-food delivery is convenient, and you like the taste—sometimes you order food twice a day. However, you realize you can't really afford this in the long term.

- Tell the nurse about your eating habits. You are set in your ways.
- Initially resist suggestions to make changes in your diet. When appropriate, acknowledge your financial worries.
- Ask for ideas about how to make gradual changes, perhaps even about easy dishes to make for yourself.

Discussion questions

1. What aspects of communicative competence (e.g., language knowledge and strategic competence) would be elicited by this task? What aspects of communicative competence would this task *not* be able to elicit?
2. How would you determine an acceptable boundary between language and content in evaluating the test taker's communicative competence on the basis of performance on this task?
3. Do you think a subject-matter expert (i.e., a practicing nurse) would judge performance differently from a professional English-language examiner? What differences would you predict in their perspectives? Whose perspective would be "correct"?

4. What elements of “test-taking communicative competence” would be required to successfully take part in the test situation itself? How would you explain the nature of this test-taking communicative competence to a student who is preparing to take this exam?
5. How could you design a rubric to capture aspects of communicative competence that rewards/does not penalize features such as accommodation, negotiation of meaning, and translanguaging as contextually appropriate?
6. What would change if this role-play was conducted online (e.g., through videoconferencing)? What else would differ if the role-play was conducted through an instant messaging/chat app?

Now consider the following discourse completion task (DCT), which could be used in an English-language proficiency test for nurses. Read through the task, and then discuss questions 1 to 4 from the prior task with reference to this task. Consider also how you might score this task fairly in operational assessment conditions.

Example task B

Discourse completion task

You are a nurse working at a general practice surgery (clinic). You are in a consultation with a 65-year-old patient who you have not met before. The patient recently broke his leg, and you need to persuade him to change his eating habits to achieve a healthy weight. The patient has already told you that he doesn't like salad or vegetables and that he prefers not to cook for himself. What would you say? (Write your response below.)

Suggestions for further reading

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Elder, C. (Ed.). (2016). Authenticity in LSP testing [Special issue]. *Language Testing*, 33(2).
- McNamara, T. F. (1996). *Measuring second language performance*. Longman. Chapter 3.
- Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal*, 95(3), 418-429.

References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Blommaert, J., Collins, J., & Slembrouck, S. (2005). Spaces of multilingualism. *Language & Communication*, 25(3), 197-216. <https://doi.org/10.1016/j.langcom.2005.05.002>
- Brindley, G. (1994). Task-centred assessment in language learning: The promise and the challenge. In N. Bird, P. Falvey, A. Tsui, D. Allison, & A. McNeill (Eds.), *Language and learning: Papers presented at the Annual International Language in Education Conference* (Hong Kong, 1993) (pp. 73-94). Hong Kong Education Department.
- Broadfoot, P. (1996). *Education, assessment and society*. Open University Press.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20(1), 1-25. <https://doi.org/10.1191/0265532203lt242oa>
- Brunfaut, T., & Harding, L. (2020). International language proficiency standards in the local context: Interpreting the CEFR in standard setting for exam reform in Luxembourg. *Assessment in Education: Principles, Policy & Practice*, 27(2), 215-231. <https://doi.org/10.1080/0969594X.2019.1700213>
- Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly*, 3(3), 229-242. https://doi.org/10.1207/s15434311laq0303_1
- Canagarajah, S. (2018). Translingual practice as spatial repertoires: Expanding the paradigm beyond structuralist orientations. *Applied Linguistics*, 39(1), 31-54. <https://doi.org/10.1093/applin/amx041>
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 2-27). London Group Ltd.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching. *Applied Linguistics*, 1(1), 1-47. <https://doi.org/10.1093/applin/I.1.1>
- Celce-Murcia, M. (2008). Rethinking the role of communicative competence in language teaching. In E. A. Soler & M. P. S. Jordà (Eds.), *Intercultural language use and language learning* (pp. 41-57). Springer.
- Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing*, 20(4), 369-383. <https://doi.org/10.1191/0265532203lt264oa>
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. M. I. T. Press.
- Council of Europe. (2001). *The common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Council of Europe. (2020). *The common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Deygers, B., Zeidler, B., Vilcu, D., & Carlsen, C. H. (2018). One framework to unite them all? Use of the CEFR in European university entrance policies. *Language Assessment Quarterly*, 15(1), 3-15. <https://doi.org/10.1080/15434303.2016.1261350>
- Deygers, B., Bigelow, M., Lo Bianco, J., Nadarajan, D., & Tani, M. (2021). Low print literacy and its representation in research and policy. *Language Assessment Quarterly*, 18(5), 463-476. <https://doi.org/10.1080/15434303.2021.1903471>
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge University Press.

- Elder, C. (2001). Assessing the language proficiency of teachers: Are there any border controls? *Language Testing*, 18(2), 149-170.
<https://doi.org/10.1177/026553220101800203>
- Elder, C., McNamara, T., Kim, H., Pill, J., & Sato, T. (2017). Interrogating the construct of communicative competence in language assessment contexts: What the non-language specialist can tell us. *Language & Communication*, 57, 14-21.
<https://doi.org/10.1016/j.langcom.2016.12.005>
- Fulcher, G. (2000). The “communicative” legacy in language testing. *System*, 28(4), 483-497.
[https://doi.org/10.1016/S0346-251X\(00\)00033-6](https://doi.org/10.1016/S0346-251X(00)00033-6)
- Fulcher, G. (2003). *Testing second language speaking*. Routledge.
- Gablasova, D., Brezina, V., Mcenery, T., & Boyd, E. (2017). Epistemic stance in spoken L2 English: The effect of task and speaker style. *Applied Linguistics*, 38(5), 613-637.
<https://doi.org/10.1093/applin/amv055>
- Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, 15(3), 219-236. <https://doi.org/10.1080/15434303.2018.1453816>
- García, O., Flores, N., Seltzer, K., Li, W., Otheguy, R., & Rosa, J. (2021). Rejecting abyssal thinking in the language and education of racialized bilinguals: A manifesto. *Critical Inquiry in Language Studies*, 18(3), 203-228.
<https://doi.org/10.1080/15427587.2021.1935957>
- Hall, J. K., Cheng, A., & Carlson, M. T. (2006). Reconceptualizing multicompetence as a theory of language knowledge. *Applied Linguistics*, 27(2), 220-240.
<https://doi.org/10.1093/applin/aml013>
- Harding, L. (2014). Communicative language testing: Current issues and future research. *Language Assessment Quarterly*, 11(2), 186-197.
<https://doi.org/10.1080/15434303.2014.895829>
- Harding, L., & Fulcher, G. (2022). Epilogue – Language testing: Where are we heading? In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed.) (pp. 622-632). Routledge. <https://doi.org/10.4324/9781003220756-47>
- Harding, L., & McNamara, T. (2018). Language assessment: The challenge of ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 570-582). Routledge.
- He, L., & Dai, Y. (2006). A corpus-based investigation into the validity of the CET-SET group discussion. *Language Testing*, 23(3), 370-401.
<https://doi.org/10.1191/0265532206lt333oa>
- Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269-293). Penguin.
- Iliovits, M., Harding, L., & Pill, J. (2022). Language use in an English-medium instruction university in Lebanon: Implications for the validity of international and local English tests for admissions. *Journal of English-Medium Instruction*, 1(2), 153-179.
<https://doi.org/10.1075/jemi.21009.ili>
- Isaacs, T. (2018). Fully automated speaking assessment: Changes to proficiency testing and the role of pronunciation. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 570-584). Routledge. <https://doi.org/10.4324/9781003220756-19>
- Knoch, U. (2022). Assessing writing. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed.) (pp. 236-253). Routledge. <https://doi.org/10.4324/9781003220756-19>
- Knoch, U., & Macqueen, S. (2020). *Assessing English for professional purposes*. Routledge.

- Kramsch, C. (2006). From communicative competence to symbolic competence. *The Modern Language Journal*, 90(2), 249-252. <https://www.jstor.org/stable/3876875>
- Kramsch, C., & Whiteside, A. (2008). Language ecology in multilingual settings. Towards a theory of symbolic competence. *Applied Linguistics*, 29(4), 645-671. <https://doi.org/10.1093/applin/amn022>
- Kunnan, A. J. (2018). *Evaluating language assessments*. Routledge.
- Leung, C. (2005). Convivial communication: Recontextualizing communicative competence. *International Journal of Applied Linguistics*, 15(2), 119-144. <https://doi.org/10.1111/j.1473-4192.2005.00084.x>
- Leung, C., & Lewkowitz, J. (2013). Language communication and communicative competence: A view from contemporary classrooms. *Language and Education*, 27(5), 398-414. <https://doi.org/10.1080/09500782.2012.707658>
- Macqueen, S. (2022). Construct in assessments of spoken language. In T. Haug, W. Mann, & U. Knoch (Eds.), *Handbook of language assessment across modalities* (pp. 233-249). Oxford University Press. <https://doi.org/10.1093/oso/9780190885052.003.0020>
- Macqueen, S., Knoch, U., Wigglesworth, G., Nordlinger, R., Singer, R., McNamara, T., & Brickle, R. (2019). The impact of national standardized literacy and numeracy testing on children and teaching staff in remote Australian Indigenous communities. *Language Testing*, 36(2), 265-287. <https://doi.org/10.1177/0265532218775758>
- Macqueen, S., Pill, J., & Knoch, U. (2021). Trust the test: Score-user perspectives on the roles of language tests in professional registration and skilled migration. *Papers in Language Testing and Assessment*, 10(1), 49-69.
- Maddox, B., Zumbo, B. D., Tay-Lim, B., & Qu, D. I. (2015). An anthropologist among the psychometricians: Assessment events, ethnography, and differential item functioning in the Mongolian Gobi. *International Journal of Testing*, 15(4), 291-309. <https://doi.org/10.1080/15305058.2015.1017103>
- Makoni, S., & Pennycook, A. (2007). *Disinventing and reconstituting languages*. Multilingual Matters.
- May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly*, 8(2), 127-145. <https://doi.org/10.1080/15434303.2011.565845>
- McNamara, T. F. (1996). *Measuring second language performance*. Longman.
- McNamara, T. F. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics*, 18(4), 446-466. <https://doi.org/10.1093/applin/18.4.446>
- McNamara, T. (2000). *Language testing*. Oxford University Press.
- McNamara, T. (2007). Language testing: A question of context. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 131-137). University of Ottawa Press.
- McNamara, T. (2011). Managing learning: Authority and language assessment. *Language Teaching*, 44(4), 500-515. <https://doi.org/10.1017/S0261444811000073>
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell.
- Milroy, J., & Milroy, L. (2012). *Authority in language: Investigating standard English* (4th ed.). Routledge.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, 19(4), 477-496. <https://doi.org/10.1191/0265532202lt241oa>
- Morrow, K. (1979). Communicative language testing: Revolution or evolution? In C. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp. 143-159). Oxford University Press.

- Morrow, K. (2012). Communicative language testing. In B. O'Sullivan & S. Stoyhoff (Eds.), *Cambridge guide to second language assessment* (pp. 140-146). Cambridge University Press.
- Motteram, J. (2020, October). *Standard English as spoken by Singaporeans: Exploring language test localization and the impact of accent in listening comprehension*. Paper presented at New Directions East Asia 2020, Singapore.
- Nakatsuhara, F., Khabbazbashi, N., & Inoue, C. (2022). Assessing speaking. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed.) (pp. 209-222). Routledge. <https://doi.org/10.4324/9781003220756-17>
- Norris, J. (2016). Current uses for task-based language assessment. *Annual Review of Applied Linguistics*, 36, 230-244. <https://doi.org/10.1017/S0267190516000027>
- Ockey, G. J., & Wagner, E. (2018). *Assessing L2 listening: Moving towards authenticity*. John Benjamins.
- O'Hagan, S., Pill, J., & Zhang, Y. (2016). Extending the scope of speaking assessment criteria in a specific-purpose language test: Operationalizing a health professional perspective. *Language Testing*, 33(2), 195-216. <https://doi.org/10.1177/0265532215607920>
- Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, 26(4), 561-587. <https://doi.org/10.1177/0265532209340192>
- Purpura, J. E. (2017). Assessing meaning. In E. Shohamy, I. G. Or, & S. May (Eds.), *Encyclopedia of language and education, Vol. 7. Language testing and assessment* (pp. 33-61). Springer International Publishing.
- Pym, H. (2017, August 22). *Do I have to understand jam-making to be a nurse?* BBC News. <https://www.bbc.co.uk/news/health-41010021>
- Roever, C., & Ikeda, N. (2022). What scores from monologic speaking tests can (not) tell us about interactional competence. *Language Testing*, 39(1), 7-29. <https://doi.org/10.1177/02655322211003332>
- Roever, C., & Kasper, G. (2018). Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing*, 35(3), 331-355. <https://doi.org/10.1177/0265532218758128>
- Ross, S. (2018). Listener response as a facet of interactional competence. *Language Testing*, 35(3), 357-375. <https://doi.org/10.1177/0265532218758125>
- Rydell, M. (2015). Performance and ideology in speaking tests for adult migrants. *Journal of Sociolinguistics*, 19(4), 535-558. <https://doi.org/10.1111/josl.12152>
- Saville-Troike, M. (1982). *The ethnography of communication*. Blackwell Publishing.
- Schissel, J. L., Leung, C., & Chalhoub-Deville, M. (2019). The construct of multilingualism in language testing. *Language Assessment Quarterly*, 16(4-5), 373-378. <https://doi.org/10.1080/15434303.2019.1680679>
- Schissel, J. L., López-Gopar, M., Leung, C., Morales, J., & Davis, J. R. (2019). Classroom-based assessments in linguistically diverse communities: A case for collaborative research methodologies. *Language Assessment Quarterly*, 16(4-5), 393-407. <https://doi.org/10.1080/15434303.2019.1678041>
- Shohamy, E. (2001a). Democratic assessment as an alternative. *Language Testing*, 18(4), 373-391. <https://doi.org/10.1177/026553220101800404>
- Shohamy, E. (2001b). *The power of tests: A critical perspective on the uses of language tests*. Longman.
- Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal*, 95(3), 418-429. <https://doi.org/10.1111/j.1540-4781.2011.01210.x>

- Spolsky, B. (1989). Communicative competence, language proficiency, and beyond. *Applied Linguistics*, 10(2), 138-156. <https://doi.org/10.1093/applin/10.2.138>
- Staples, S., LaFlair, G. T., & Egbert, J. (2017). Comparing language use in oral proficiency interviews to target domains: Conversational, academic, and professional discourse. *The Modern Language Journal*, 101(1), 194-213. <https://doi.org/10.1111/modl.12385>
- Toomaneejinda, A. (2018). *An exploratory study of expressing disagreement in ELF academic group discussion* [Unpublished doctoral thesis]. Lancaster University.
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford University Press.
- Yin, C. & Mislevy, R. (2022). Evidence-centered design in language testing. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed.) (pp. 289-305). Routledge. <https://doi.org/10.4324/9781003220756-23>
- Youn, S. J. (2020). Managing proposal sequences in role-play assessment: Validity evidence of interactional competence across levels. *Language Testing*, 37(1), 76-106. <https://doi.org/10.1177/0265532219860077>
- Young, R., & He, A. W. (Eds.). (1998). *Talking and testing: Discourse approaches to the assessment of oral proficiency*. John Benjamins.