

Encyclopedia of
Language and Education
Series Editor: Stephen May

SPRINGER
REFERENC



RAHNAMA
P R E S S

@RAHNAMAPRESS
WWW.RAHNAMAPRESS.COM

Elana Shohamy · Iair G. Or
Stephen May *Editors*

Language Testing and Assessment

Third Edition

 Springer

Editor in Chief's Introduction to the "Encyclopedia of Language and Education"

This is one of ten volumes of the *Encyclopedia of Language and Education* published by Springer. The *Encyclopedia* – now in this, its third edition – is undoubtedly the benchmark reference text in its field. It was first published in 1997 under the general editorship of the late David Corson and comprised eight volumes, each focused on a single, substantive topic in language and education. These included: language policy and political issues in education; literacy; oral discourse and education; second language education; bilingual education; knowledge about language; language testing and assessment; and research methods in language and education.

In his introductory remarks, David made the case for the timeliness of an overarching, state-of-the-art review of the language and education field. He argued that the publication of the *Encyclopedia* reflected both the internationalism and interdisciplinarity of those engaged in the academic analysis of language and education, confirmed the maturity and cohesion of the field, and highlighted the significance of the questions addressed within its remit. Contributors across the first edition's eight volumes came from every continent and from over 40 countries. This perhaps explains the subsequent impact and reach of that first edition – although no one (except, perhaps, the publisher!) quite predicted its extent. The *Encyclopedia* was awarded a Choice Outstanding Academic Title Award by the American Library Association and was read widely by scholars and students alike around the globe.

In 2008, the second edition of the *Encyclopedia* was published under the general editorship of Nancy Hornberger. It grew to ten volumes as Nancy continued to build upon the reach and influence of the *Encyclopedia*. A particular priority in the second edition was the continued expansion of contributing scholars from contexts outside of English-speaking and/or developed contexts, as well as the more effective thematic integration of their regional concerns across the *Encyclopedia* as a whole. The second edition also foregrounded key developments in the language and education field over the previous decade, introducing two new volumes on language socialization and language ecology.

This third edition continues both the legacy and significance of the previous editions of the *Encyclopedia*. A further decade on, it consolidates, reflects, and expands (upon) the key issues in the field of language education. As with its predecessors, it overviews in substantive contributions of approximately 5000

words each, the historical development, current developments and challenges, future directions, of a wide range of topics in language and education. The geographical focus and location of its authors, all chosen as experts in their respective topic areas, also continues to expand, as the *Encyclopedia* aims to provide the most representative international overview of the field to date.

To this end, some additional changes have been made. The emergence over the last decade of "superdiversity" as a topic of major concern in sociolinguistics, applied linguistics, and language education is now a major thread across all volumes – exploring the implications for language and education of rapidly changing processes of migration and transmigration in this late capitalist, globalized world. This interest in superdiversity foregrounds the burgeoning and rapidly complexifying uses of language(s), along with their concomitant deconstruction and (re)modification, across the globe, particularly (but not exclusively) in large urban environments. The allied emergence of multilingualism as an essential area of study – challenging the long-held normative ascendancy of monolingualism in relation to language acquisition, use, teaching, and learning – is similarly highlighted throughout all ten volumes, as are their pedagogical consequences (most notably, perhaps, in relation to translanguaging). This "multilingual turn" is reflected, in particular, in changes in title to two existing volumes: *Bilingual and Multilingual Education* and *Language Awareness, Bilingualism and Multilingualism* (previously, *Bilingual Education* and *Language Awareness*, respectively).

As for the composition of the volumes, while ten volumes remain overall, the *Language Ecology* volume in the 2nd edition was not included in the current edition, although many of its chapter contributions have been reincorporated and/or reworked across other volumes, particularly in light of the more recent developments in superdiversity and multilingualism, as just outlined. (And, of course, the important contribution of the *Language Ecology* volume, with Angela Creese and the late Peter Martin as principal editors, remains available as part of the second edition.) Instead, this current edition has included a new volume on *Language, Education and Technology*, with Steven Thorne as principal editor. While widely discussed across the various volumes in the second edition, the prominence and rapidity of developments over the last decade in academic discussions that address technology, new media, virtual environments, and multimodality, along with their wider social and educational implications, simply demanded a dedicated volume.

And speaking of multimodality, a new, essential feature of the current edition of the *Encyclopedia* is its multiplatform format. You can access individual chapters from any volume electronically, you can read individual volumes electronically and/or in print, and, of course, for libraries, the ten volumes of the *Encyclopedia* still constitute an indispensable overarching electronic and/or print resource.

As you might expect, bringing together ten volumes and over 325 individual chapter contributions has been a monumental task, which began for me at least in 2013 when, at Nancy Hornberger's invitation, Springer first approached me about the Editor-in-Chief role. All that has been accomplished since would simply not have occurred, however, without support from a range of key sources. First, to Nancy Hornberger, who, having somehow convinced me to take on the role, graciously

agreed to be Consulting Editor for the third edition of the *Encyclopedia*, providing advice, guidance, and review support throughout.

The international and interdisciplinary strengths of the *Encyclopedia* continue to be foregrounded in the wider topic and review expertise of its editorial advisory board, with several members having had direct associations with previous editions of the *Encyclopedia* in various capacities. My thanks to Suresh Canagarajah, William Cope, Viv Edwards, Rainer Enrique Hamel, Eli Hinkel, Francis Hult, Nkonko Kamwangamalu, Gregory Kamwendo, Claire Kramersch, Constant Leung, Li Wei, Luis Enrique Lopez, Marilyn Martin-Jones, Bonny Norton, Tope Omoniyi, Alastair Pennycook, Bernard Spolsky, Lionel Wee, and Jane Zuengler for their academic and collegial support here.

The role of volume editor is, of course, a central one in shaping, updating, revising, and, in some cases, resituating specific topic areas. The third edition of the *Encyclopedia* is a mix of existing volume editors from the previous edition (Cenoz, Duff, King, Shohamy, Street, Van Deusen-Scholl), new principal volume editors (García, Kim, Lin, McCarty, Thorne, Wortham), and new coeditors (Lai, Or). As principal editor of *Language Policy and Political Issues in Education*, Teresa McCarty brings to the volume her longstanding interests in language policy, language education, and linguistic anthropology, arising from her work in Native American language education and Indigenous education internationally. For *Literacies and Language Education*, Brian Street brings a background in social and cultural anthropology, and critical literacy, drawing on his work in Britain, Iran, and around the globe. As principal editors of *Discourse and Education*, Stanton Wortham has research expertise in discourse analysis, linguistic anthropology, identity and learning, narrative self-construction, and the new Latino diaspora, while Deoksoon Kim's research has focused on language learning and literacy education, and instructional technology in second language learning and teacher education. For *Second and Foreign Language Education*, Nelleke Van Deusen-Scholl has academic interests in linguistics and sociolinguistics and has worked primarily in the Netherlands and the United States. As principal editors of *Bilingual and Multilingual Education*, Ofelia García and Angel Lin bring to the volume their internationally recognized expertise in bilingual and multilingual education, including their pioneering contributions to translanguaging, along with their own work in North America and Southeast Asia. Jasone Cenoz and Durk Gorter, principal editors of *Language Awareness, Bilingualism and Multilingualism*, bring to their volume their international expertise in language awareness, bilingual and multilingual education, linguistic landscape, and translanguaging, along with their work in the Basque Country and the Netherlands. Principal editor of *Language Testing and Assessment*, Elana Shohamy, is an applied linguist with interests in critical language policy, language testing and measurement, and linguistic landscape research, with her own work focused primarily on Israel and the United States. For *Language Socialization*, Patricia Duff has interests in applied linguistics and sociolinguistics and has worked primarily in North America, East Asia, and Central Europe. For *Language, Education and Technology*, Steven Thorne's research interests include second language acquisition, new media and online gaming environments, and

theoretical and empirical investigations of language, interactivity, and development with his work focused primarily in the United States and Europe. And for *Research Methods in Language and Education*, principal editor, Kendall King, has research interests in sociolinguistics and educational linguistics, particularly with respect to Indigenous language education, with work in Ecuador, Sweden, and the United States. Finally, as Editor-in-Chief, I bring my interdisciplinary background in the sociology of language, sociolinguistics, applied linguistics, and educational linguistics, with particular interests in language policy, Indigenous language education, and bilingual education, along with my own work in New Zealand, North America, and the UK/Europe.

In addition to the above, my thanks go to Yi-Ju Lai, coeditor with Kendall King, and Lair G. Or, coeditor with Elana Shohamy. Also to Lincoln Dam, who as Editorial Assistant was an essential support to me as Editor-in-Chief and who worked closely with volume editors and Springer staff throughout the process to ensure both its timeliness and its smooth functioning (at least, to the degree possible, given the complexities involved in this multiyear project). And, of course, my thanks too to the approximately 400 chapter contributors, who have provided the substantive content across the ten volumes of the *Encyclopedia* and who hail from every continent in the world and from over 50 countries.

What this all indicates is that the *Encyclopedia* is, without doubt, not only a major academic endeavor, dependent on the academic expertise and goodwill of all its contributors, but also still demonstrably at the cutting edge of developments in the field of language and education. It is an essential reference for every university and college library around the world that serves a faculty or school of education and is an important allied reference for those working in applied linguistics and sociolinguistics. The *Encyclopedia* also continues to aim to speak to a prospective readership that is avowedly multinational and to do so as unambiguously as possible. Its ten volumes highlight its comprehensiveness, while the individual volumes provide the discrete, in-depth analysis necessary for exploring specific topic areas. These state-of-the-art volumes also thus offer highly authoritative course textbooks in the areas suggested by their titles.

This third edition of the *Encyclopedia of Language and Education* continues to showcase the central role of language as both vehicle and mediator of educational processes, along with the pedagogical implications therein. This is all the more important, given the rapid demographic and technological changes we face in this increasingly globalized world and, inevitably, by extension, in education. But the cutting-edge contributions within this *Encyclopedia* also, crucially, always situate these developments within their historical context, providing a necessary *diachronic* analytical framework with which to examine *critically* the language and education field. Maintaining this sense of historicity and critical reflexivity, while embracing the latest developments in our field, is indeed precisely what sets this *Encyclopedia* apart.

The University of Auckland
Auckland, New Zealand

Stephen May

Volume Editors' Introduction to "Language Testing and Assessment"

This volume addresses the broad theme and specific topics associated with current thinking in the field of language testing and assessment. Interdisciplinary in their nature, language testing and assessment build on theories and definitions provided by linguistics, applied linguistics, language acquisition, and language teaching, as well as on the disciplines of testing, measurement, and evaluation. Language testing uses these disciplines as foundations for researching, theorizing, and constructing valid language tools for assessing and judging the quality of language. Language testing and assessment are always historically situated and conditioned, embedded in knowledge, beliefs, and ideologies about their goals and best practices. They also play an important role in education, policy, and society, and their educational and societal consequences cannot be ignored. The present volume therefore responds to the high demand for clear, reliable, and up-to-date information about language testing and assessment theories and practices, while keeping in sight the rich social contexts in which they function.

The main focus of this volume, which sets it apart from similar volumes and handbooks, is innovation. We wanted the volume to present state-of-the-art techniques, principles, insights, and methodologies for a new generation of practitioners, researchers, and experts in language testing and assessment. For this purpose, we selected a range of topics which, while providing a broad overview of the field, focuses on advances and breakthroughs of the past decade or so. As a consequence, many of the topics in this volume – such as multilingual assessment, the assessment of meaning, English as a lingua franca (ELF), the Common European Framework of Reference (CEFR), the Common Core policy in the USA, or critical testing – are covered for the first time in a volume of this sort by experts dedicated to them. Of the volume's 29 chapters, 15 are completely new, many of them covering aspects of language assessment that were not included in the second edition of this encyclopedia, published in 2008. In addition to that, we uniformly asked all the authors – both those contributing to the volume for the first time and those updating their contributions from the previous edition – to report about innovations, new research, or novel techniques in their area of expertise. Consequently, this third edition volume can be seen as groundbreaking, strongly emphasizing recent developments, as well as providing an outlook of the future of this dynamic field.

The field of language testing is traditionally viewed as consisting of two main components: one focusing on the “what,” referring to the constructs that need to be assessed (also known as “the trait”), and the other component pertaining to the “how” (also known as “the method”), which addresses the specific procedures and strategies used for assessing the “what.” Traditionally, “the trait” has been defined by the language testing field; these definitions have provided the essential elements for creating language tests. The “how,” on the other hand, is derived mostly from the field of testing and assessment which has, over the years, developed a broad body of theories, research, techniques, and practices. Today, a crucial third component is added to the field, focusing on language assessment practices and the social consequences and implications of language testing and assessment. Language testers incorporate these three areas to create the discipline of language testing and assessment, a field which includes theories, research, and applications; it has its own research publications, conferences, and two major journals, *Language Testing* and *Language Assessment Quarterly*, where many of these studies appear.

An examination of the developments in the language testing and assessment field since the 1960s reveals that its theories and practices have always been closely related to definitions of language proficiency. Matching the “how” of testing with the “what” of language uncovers several periods in the development of the field, with each one instantiating different notions of language knowledge along with specific measurement procedures that go with them. Thus, discrete-point testing viewed language as consisting of lexical and structural items so that the language test of that era presented isolated items in objective testing procedures. In the integrative era, language tests tapped integrated and discursal language; in the communicative era, tests aimed to replicate interactions among language users utilizing authentic oral and written texts; and in the performance testing era, language users were expected to perform tasks taken from “real life” contexts. Alternative assessment was a way of responding to the realization that language knowledge is a complex phenomenon, which no single procedure can be expected to capture. Assessing language knowledge therefore requires multiple and varied procedures that complement one another. While we have come to accept the centrality of the “what” to the “how” trajectory for the development of tests and assessment instruments, extensive work in the past two decades has pointed to a less overt but highly influential dynamic in another direction. This dynamic has to do with the pivotal roles that tests play in societies in shaping the definitions of language, in affecting learning and teaching, and in maintaining and creating social classes. This means that contemporary assessment research perceives as part of its obligations the need to examine the close relationship between methods and traits in broader contexts and to focus on how language tests interact with societal factors, given their enormous power. In other words, as language testers seek to develop and design methods and procedures for assessment (the “how”) they become mindful not only of the emerging insights regarding the trait (the “what”), and its multiple facets and dimensions, but also of the societal role that language tests play, the power that they hold, and their central functions in education, politics, and society.

In terms of the interaction of society and language, it is evident that change is currently occurring in the broader contexts and spaces in which language assessment takes place. It is increasingly realized nowadays that language assessment does not occur in homogeneous, uniform, and isolated contexts but, rather, in diverse, multilingual, and multicultural societies. This in turn poses new challenges and questions with regards to what it means to know language(s) in education and society. For example, different meanings of language knowledge may be associated with learning foreign languages, second languages, language by immersion, heritage languages, languages of immigrants arriving to new places with no knowledge of the new languages, multilingualism and translanguaging practices by those defined as "transnationals," and English as a lingua franca, for which language knowledge is different from the knowledge of other languages. As a consequence, the current focus on multilingualism, translanguaging, lingua franca, immigrants/refugees/asylum seekers, etc. has been incorporated in many of the chapters of this volume.

Similarly, the language of classrooms and schools may be different from that of the workplaces or communities where bi- or multilingual patterns are the norm. Each of these contexts may require different and varied theories of language knowledge and hence different definitions, applications, and methods of measuring these proficiencies. In other words, the languages currently used in different societies and in different contexts no longer represent uniform constructs, as these vary from one place to another, from one context to another, creating different language patterns, expectations, and goals, and often resulting in linguistic hybrids and fusions. Such dynamic linguistic phenomena pose challenges for language testers. What is the language (or languages) that needs to be assessed? Where can it be observed in the best ways? Is it different at home, in schools, in classrooms, and in the workplace? Should hybrids and fusions be assessed and how? Should multilingual proficiencies be assessed and how? Can levels of languages even be defined? How should language proficiency be reported and to whom? What is "good language"? Does such a term even apply? Who should decide how tests should be used? Do testers have an obligation to express their views about language and testing policy? What is the responsibility of testers to language learning and language use in classrooms and communities? How can ethical and professional attitudes in the field be maintained? These are some of the questions with which language testers are currently preoccupied. Language testers are not technicians that just invent better and more sophisticated testing tools. Rather, they are constantly in search for and concerned with the "what" and its complex meanings. Going beyond general testing, the unique aspect of language testing is that it is an integral part of a defined discipline, that of "Language." In this respect, language testers and the field of language testing and assessment are different from the field of general testing in that language testers are confined to a specific discipline and are therefore in constant need of asking such language-related questions as listed above in order to develop valid language assessment tools. Yet, even this list of questions is changing and context-dependent, since language today cannot be detached from multiple social, cultural, linguistic, and political dynamics.

The concern of language testers in the past two decades about the use of test their political, social, educational, and ethical dimensions has made the field more complex and uncertain and in need of new discussions and debates. Elana Shohamy, the editor of this volume in the 2008 edition, stated that the era we are in could be described as the era of uncertainty, where questions are being raised about the meaning of language, along with the possibilities for measuring this complex and dynamic variable. While this statement still holds true, we may be experiencing times where some (complex, initial) answers and solutions for some of these questions are beginning to emerge. We are in an era where there is an ever more compelling need to ensure that these tests are reliable and valid, where validity includes the protection of the personal rights of others, as well as positive washback on learning by addressing the diverse communities in which the tests are used. Thus, the current era is not only concerned with a broader and more complex view of what it means to know a language, or with innovative methods of testing and assessment of complex constructs, but also with how these tests can be more inclusive, democratic, just, open, fair and equal, and less biased. Even within the use of traditional large-scale testing, the field is asking questions about test use: Why test? Who benefits, who loses? What is the impact on and consequences for definitions of language in relation to people, education, language policy, and society? Tests are no longer viewed as innocent tools, but rather as instruments that play central roles for people, education, and societies. Language testers, therefore, are asked to deal with and find solutions to broader issues: to examine the uses of tests in the complex multilingual and multicultural societies where they are used, not only as naïve measurement tools but also as powerful educational, societal, and political devices. This is the conceptual premise of this third edition volume of the *Encyclopedia of Language and Education on Language and Assessment*. It aims to cover (and uncover) the multiple versions and perspectives of the “what” of languages along with the multiple approaches developed for assessment of the “what,” especially given the multiplicity of languages used by many diverse groups of learners in many different contexts. It aims to focus on the societal roles of language testers and their responsibility to be socially accountable and to ensure ethicality and professionalism. It also strives to show some of the emerging solutions and new directions that try to address these issues. A special focus is given in this volume to the multilingual and diverse contexts in which language testing and assessment are currently anchored and the difficult task of language testing and assessment in this complex day and age.

Accordingly, the first part of the volume addresses the “what” of language testing and assessment, looking into the constructs and domains of language assessment. Rather than dividing language into neat and clear-cut skills of reading, writing, speaking, and listening, it examines the “what” of language in the diverse contexts in which it is used. Instead of proposing one uniform way of defining the language construct, the chapters in Part 1 present language from multiple perspectives, which represent a variety of language activities. It begins with Lorena Llosa’s chapter on the assessment of students’ content knowledge and language proficiency, showing the complex, dynamic relations between content knowledge and language, critiquing

the traditional separation between the two and discussing recent attempts to integrate them in assessment. In the next chapter, Angela Scarino explores the position and role of culture in language assessment in times of increased globalization, multilinguality, and multiculturalism. She argues that the construct of culture is and should be reconsidered to reflect complex realities, challenging established language assessment paradigms and raising ethical issues. James Purpura, in a novel contribution for such a volume, explores the construct of meaning and remaps the history of language testing through the lens of meaning-making. He shows that the focus since the 1980s on functional proficiency has been at the cost of meaning-making and propositional content and suggests various paths for assessing meaning. Rachel Brooks examines the changing language assessment practices and norms in the US government, as a large-scale example of language assessment at the workplace. Consisting of a wide range of departments, organizations, and aims, government activity greatly relies on high-stakes language testing, and some of its agencies are also involved in language testing development and research. Megan Smith and Charles Stansfield's chapter focuses on the language aptitude construct and the role of language aptitude tests in second language learning. The authors track the developments in the theory and practice of language aptitude measurement, as well as recent attempts to validate or find alternatives to the ways in which language aptitude is measured.

The concluding two chapters of the first part focus on recent challenges and innovations that represent two growing fields of language assessment. In their chapter on the assessment of multilingual competence, Alexis Augusto López, Sultan Turkan, and Danielle Guzman-Orth discuss the growing recognition, even by large testing authorities, that multilingual assessment tools are necessary for validly measuring the language knowledge of multilinguals in contexts of immigration or complex, globalized language realities. Although the field of multilingual assessment is still nascent, the authors present some of the early attempts that have already been made and discuss their importance and characteristics. Similarly, the chapter by Jennifer Jenkins on the assessment of English as a lingua franca (ELF) presents a field that seeks to answer the needs of globalized, transnational, "super-diverse" societies, in which English plays a major role as the shared language of non-native English speakers. Although no implementations of ELF tests and assessments have been developed so far, Jenkins outlines the goals, constructs, and limitations of such prospective tests, thereby proposing a novel outlook on how language testing can become more closely linked to the ways in which English is actually used as a second or foreign language. Together, these seven chapters provide multiple perspectives of the language constructs and assessment practices associated with them. As these chapters show, definitions of language cannot be detached from the diverse contexts in which they are used.

The second part of the volume addresses the methodological issues that language testers face when assessing the complex construct of language: that is, the "how." The chapters explore a wide variety of approaches and procedures for assessing language, each with its theoretical underpinnings and motivations and the issues it addresses. In the first chapter, Gillian Wigglesworth and Kellie Frost survey task and performance-based assessment, among the most popular alternative assessment tools

today, designed to measure learners' productive and receptive language through performances related to real world contexts. They discuss the value of certain performance tests, the extent to which they indeed represent "real life," and the recent trend of moving away from individual components of language proficiency to integrated tasks incorporating more than one skill. Staying within the context of alternative assessment, Janna Fox provides an overview of the various techniques, focusing on portfolio assessment, which has become the most pervasive approach. She discusses the usefulness of portfolios for both formative and summative assessment, as well as their claim for authenticity. Finally, she reviews the impact of newer technologies in the development of e-portfolios and other forms of digital learner records.

The implications of technology for language assessment are the topic of the next chapter, written by Carol Chapelle and Erik Voss, who begin their chapter with a historical overview of computer-assisted language testing, showing how technological advancements led to the development of computer-adaptive testing and natural language processing techniques. The authors discuss the potential influence of technology on test performance as part of the current and future challenges in the field. The chapter by Eunice Jang traces the cognitive processes involved in language assessment, looking into learner cognition and the way assessment tools should be devised to address various processes and their dynamic interplay with learners' multiple traits. Jang concludes the chapter by pointing to some future possibilities of harnessing technology to make assessment processes less intrusive. Glenn Fulcher provides a comprehensive description of the methods used for examining the quality of language via rating scales, standards, benchmarks, band levels, frameworks, and guidelines. He shows the advantages and disadvantages of these tools in terms of validity of progression, equivalence across languages, hierarchies, and misconceptions serving as criteria for language assessment. He stresses the fact that psychometrics has gone through major changes and has been replaced with a more pluralistic philosophical environment, in which consensus about language quality criteria no longer exists.

The chapter by Xiaoming Xi and Yasuyo Sawaki explores quantitative and qualitative methods of test validation, examining the evolution of validity theory and validation frameworks in general and argument-based validation in particular, and the issues associated with it. The authors also discuss the emergence of alternative validation approaches, constantly challenged by new concepts and constructs such as English as a lingua franca, new technologies, and new language learning frameworks. In continuation with the discussion of validation, Anne Lazaraton describes in her chapter the tensions between various approaches for validation and describes the increasingly popular qualitative approaches and techniques used for designing and evaluating performance tests. She surveys some of the key studies in this field, showing the merits of a mixed-methods approach, and discusses the main challenges faced by qualitative validation today. Concluding this section, Meg Malone's contribution focuses on training designed to increase language assessment literacy among teachers, principals, policy makers, and other agents. She reviews the major approaches in training, affected by changes in the educational, societal, and

philosophical contexts of testing. By analyzing textbooks for language assessment, she tracks the main developments in training and outlines some of the main issues, such as the scarcity of resources and lack of agreement between language testers and teachers regarding the main building blocks of language assessment literacy.

While the chapters of the second part highlight the practices and innovations in language assessment methods, from design to validation and training, the third part of this volume looks into language assessment as it is embedded in educational systems and contexts, where language assessment and especially tests are so widely used. It is in the educational system that tests and various assessment methods serve as major tools for: assessing language for learning and teaching, making decisions about programs, teachers and learners, and finally creating changes that lead to school reforms and bring intended and unintended washback in classrooms and schools. Matthew Poehner, Kristin Davin, and James Lantolf open this part with a chapter on dynamic assessment (DA), which is one of the most promising approaches to assessment in education. DA undertakes language assessment by applying Vygotsky's sociocultural theories, closely linking assessment and learning. The authors discuss the growing body of research in the field and emphasize the effectiveness of this approach with multiple populations, including immigrants, young learners, gifted learners, and learners with special needs. They conclude by discussing current studies on computerized administration of DA. Ofra Inbar-Lourie unravels the new concept of language assessment literacy (LAL) as an umbrella term for the knowledge, skills, and background that various participants in language assessment are expected to master. She explores the history of this concept and the challenges of arriving at an agreed upon set of skills or principles shared by the entire educational community. Looking into the future of this domain, she concludes that one of the most promising areas involves the creation of situated, differential LAL rather than a unified one.

The next five chapters are devoted to specific contexts of language assessment in education. Catherine Elder analyzes language assessment in the context of higher education, which is becoming a major site of Englishization and internationalization as well as language assessment expertise. Used for a wide variety of purposes, language assessment in higher education is often driven by powerful testing agencies, which in some cases limit the ability to develop local assessment policies for diverse student populations and for the introduction of new technologies. Beverly Baker and Gillian Wigglesworth delve into the Indigenous contexts of Australia and Canada – a research focus which is gaining recognition among researchers and policy makers. Against the backdrop of the historical mistreatment of Indigenous populations, both countries pay increased attention to language assessment as part of language revitalization and bilingual education efforts. The authors present some recent evidence showing that there is a growing acknowledgment of the importance of community participation in language assessment policies. Jamal Abedi looks into another intricate context of language assessment – that of using accommodations for learners with various disabilities or impairments, as well as for language learners in immigration contexts. Reviewing the extensive research conducted in the past two decades in the topic, he examines the effectivity and validity of accommodations for

language learners, mostly in the context of English language learners in the US. He concludes with a set of principles regarding the need to limit the accommodations to the elimination of construct-irrelevant influences. Focusing on yet another language assessment context of expanding interest, Alison Bailey's chapter discusses young language learners (aged 3–11), who require a unique set of methods and techniques for assessing their language. Pointing to the different strategies of these kinds of tests compared with those used for adults, she explores the potential and limitations of the field, which is gaining major attention nowadays as it becomes ever more widely implemented. Constant Leung and Jo Lewkowicz complete this tour of language assessment contexts by surveying second or additional language assessment of linguistic minority students and in contexts where bi- or multilingualism is strongly encouraged, as in the European Union. They elucidate some of the constructs and recent developments, pointing at future directions which recognize the multiple linguistic repertoires and proficiencies of diverse populations and avoid the imposition of one language assessment standard on all.

Concluding the third part of the volume, Dina Tsagari and Liying Cheng delve into the study of the unavoidable washback, impact, and consequences assessment has on learning, teaching, and curriculum development. Tracking the long history of research into the impact and consequences of testing and distinguishing between two major strands of studies, they focus on recent studies, claiming that the complexity of these educational phenomena and the controversies surrounding them pose a serious challenge for any future study of these domains as well as for their interaction with notions of validity, fairness, and ethics in language assessment. Taken together, the chapters in Part 3 cover a wide range of topics related to broad issues of language assessment in education, especially amidst the changing realities of school demographics with regards to diverse populations and the role assessment can play in bringing about educational reform.

The fourth and final part of this volume puts language testing and assessment in a broader context, addressing the societal, political, professional, and ethical dimensions of assessments and tests. This topic has been a major concern in the language assessment field since the 1990s, and its importance is gaining broader recognition. Each of the six chapters in this section explores a different aspect of these dimensions. The section begins with a historical survey by Bernard Spolsky, in which the past, present, and future of the field are discussed, providing guidance and direction for the future. Spolsky surveys the advances in the field as well as the ample questions, contradictions, and uncertainties that need to be addressed in the future. He ends the chapter by stating that he remains skeptical about language testing, given the role of industrial test-makers in computerizing tests and in reducing multidimensional language profiles into uniform scales, and also given that educational systems continue to interpret test scores as if they are meaningful. At the same time, he expects the quality research that has been conducted in the field of language testing to continue, especially that which has been conducted in relation to the "nature" of language proficiency and the diverse approaches to assessing it in various social contexts. The chapter by Kate Menken illustrates how high-stakes language tests represent de facto language policies that affect schools and societies

and deliver direct messages about the significance and insignificance of various languages and language instruction policies. Menken reviews the history of standardized testing and the detrimental impact of monolingual testing on education. She underlines the consequences of monolingual testing and proposes the adoption of multilingual assessment and translanguaging theory as a way to counter those problems, addressing immigrant and ELL populations.

The following chapter, on ethics, professionalism, rights, and codes, by the late Alan Davies, is included posthumously; we had the great honor of having him revise and update his contribution not long before his passing. Davies, who has written extensively on the ethical dimensions of tests and the professional aspects related to ethicality, addresses these issues by covering the developments in the language testing field, showing how the code of ethics and code of practice, developed by the language testing profession via the International Language Testing Association (ILTA), can lead to the more ethical use of tests, and questioning the effectiveness of this and similar courses of action. Davies warns against the use of ethical codes as face-saving devices, which, he argues, overlooks the real commitment to ethics that is instrumental for the profession itself, for its stakeholders, and for the rights of test-takers. He also proposes a model for the ethicality of tests for asylum seekers and the inappropriate use of tests by state authorities. This chapter is followed by two chapters that may illustrate some of the ethical complexities of language assessment, focusing on two major educational and societal contexts. First, Monica Barni and Luisa Salvati reflect on the uses and misuses of the Common European Framework (CEFR) for languages, originally designed to promote multilingualism and cultural diversity but eventually used by policy makers as a tool for the selection of migrant populations. Using the Italian situation as an example, the authors discuss the lack of reflection and consideration of the way the CEFR is used and the extent of its dangerous attraction for politicians and lawmakers, who tend to adopt it without considering the theory, know-how, and limitations of this tool from a professional point of view. Second, the chapter by Luis E. Poza and Guadalupe Valdés explores the recent history of English language assessment in the USA from the No Child Left Behind Act to the Common Core. The authors outline the tremendous impact of these two policies, which force schools and states to be constantly evaluated and particularly to develop or adopt new standards for English as a second language. The result has been the imposition of a standardizing testing-driven regime on English language learners (ELLs) who greatly vary in their levels of bilingualism and English-language proficiency. Poza and Valdés conclude by pointing at future directions that may mitigate some of the problems and improve the overall level of ESL, which is such a crucial component of education in the USA.

The concluding chapter of this volume, by Elana Shohamy, takes a critical look at testing by examining the critical issues arising from language testing in a variety of contexts. She discusses the critical language testing (CLT) research agenda proposed by her and other authors in the past two decades, focusing on the power of tests and the ways it can and should be addressed. By going back to many of the contributions in this volume, Shohamy points at various directions in which current research in the language assessment domain can tackle the issues created by the often detrimental

effects of language testing, suggesting constructive and positive forms of language assessment, enhancing equality and justice in this domain, and encompassing definitions of language that are more pertinent to our times.

The editors would like to thank each and every author of these chapters, which together make up a most valuable contribution to current thinking in the field of language testing and applied linguistics. The authors selected to write these chapters are among the most distinguished scholars and leaders in the field of language testing and assessment internationally. The chapters herein reveal that the language testing field is dynamic, thriving, and vital. It is clear from these chapters that the field of language testing raises deep, important questions and does not overlook problems, difficulties, contradictions, malpractices, and new societal realities and needs. While viewed by some as a technical field, this volume convincingly demonstrates that language testing and assessment is, above all, a scholarly and intellectual field that touches the essence of languages in their deepest meanings. The need to get engaged in testing and assessment forces testers to face these issues head-on and attempt to deliberate on creative and thoughtful solutions which benefit society and are professional and ethically responsible.

Tel Aviv

Elana Shohamy
Iair G. Or

Contents

Part I Assessing Language Domains	1
Assessing Students' Content Knowledge and Language Proficiency ...	3
Lorena Llosa	
Culture and Language Assessment	15
Angela Scarino	
Assessing Meaning	33
James Enos Purpura	
Language Assessment in the US Government	63
Rachel L. Brooks	
Testing Aptitude for Second Language Learning	77
Megan Smith and Charles W. Stansfield	
Assessing Multilingual Competence	91
Alexis A. Lopez, Sultan Turkan, and Danielle Guzman-Orth	
Assessing English as a Lingua Franca	103
Jennifer Jenkins and Constant Leung	
Part II Methods of Assessment	119
Task and Performance-Based Assessment	121
Gillian Wigglesworth and Kellie Frost	
Using Portfolios for Assessment/Alternative Assessment	135
Janna Fox	
Utilizing Technology in Language Assessment	149
Carol A. Chapelle and Erik Voss	
Cognitive Aspects of Language Assessment	163
Eunice Eunhee Jang	



Criteria for Evaluating Language Quality 193
 Glenn Fulcher

Methods of Test Validation 211
 Xiaoming Xi and Yasuyo Sawaki

Qualitative Methods of Validation 225
 Anne Lazaraton

Training in Language Assessment 241
 Margaret E. Malone

Part III Assessment in Education **241**

Dynamic Assessment 243
 Matthew E. Poehner, Kristin J. Davin, and James P. Lantolf

Language Assessment Literacy 257
 Ofra Inbar-Lourie

Language Assessment in Higher Education 271
 Catherine Elder

Language Assessment in Indigenous Contexts in Australia and Canada 287
 Beverly Baker and Gillian Wigglesworth

Utilizing Accommodations in Assessment 303
 Jamal Abedi

Assessing the Language of Young Learners 323
 Alison L. Bailey

Assessing Second/Additional Language of Diverse Populations 343
 Constant Leung and Jo Lewkowicz

Washback, Impact, and Consequences Revisited 359
 Dina Tsagari and Liying Cheng

Part IV Assessment in Society **373**

History of Language Testing 375
 Bernard Spolsky

High-Stakes Tests as De Facto Language Education Policies 385
 Kate Menken

Ethics, Professionalism, Rights, and Codes 397
 Alan Davies



The Common European Framework of Reference (CEFR)	
Monica Barni and Luisa Salvati	
Assessing English Language Proficiency in the United States	427
Luis E. Poza and Guadalupe Valdés	
Critical Language Testing	441
Elana Shohamy	
Index	455

Assessing Students' Content Knowledge and Language Proficiency

Lorena Llosa

Abstract

The relationship between language proficiency and content knowledge in assessment is a complicated one. From the perspective of content assessment, language has typically been considered a source of construct-irrelevant variance. From the perspective of language assessment, content has also been considered a potential source of construct-irrelevant variance. However, regardless of the purpose for assessment, both content knowledge and language proficiency are engaged to some extent. This chapter explores how the relationship between these two constructs has been conceptualized in the field of language assessment.

Keywords

Language assessment • Content assessment • English language learners

Contents

Introduction	4
Early Developments	5
Major Contributions	5
Language for Specific Purpose Testing	6
Content and Language Assessment of ELLs in Schools	7
Work in Progress	9
Problems and Difficulties	10
Future Directions	10
Cross-References	12
Related Articles in the Encyclopedia of Language and Education,	12
References	12

L. Llosa (✉)

Multilingual Multicultural Studies, Department of Teaching and Learning, New York University,
New York, NY, USA

e-mail: lorena.llosa@nyu.edu



Introduction

The relationship between language proficiency and content knowledge in assessment has always been a complicated one. From the perspective of content assessment, language has typically been considered a source of construct-irrelevant variance – variance in scores that is not related to the construct being assessed. From the perspective of language assessment, content (also referred to as topical knowledge or background knowledge) has also been considered a potential source of construct-irrelevant variance. Thus for the purpose of assessment, language proficiency and content knowledge have traditionally been viewed as separate and distinct constructs. The language ability models that have informed the constructs of most language assessments (e.g., Bachman and Palmer 1996) included topical knowledge as a category of language use, but one that was separate from language knowledge and strategic competence.

Regardless of the purpose of an assessment – either to assess a test taker’s language proficiency or their content knowledge in a particular area – these two constructs cannot be so easily disentangled. Any assessment of content will involve language, and any assessment of language that will be useful for making inferences about a test taker’s ability to use language in a context outside the test itself will involve some content or topical knowledge. Therefore the nature of the content-language link and the role it plays in construct definitions when assessing learners of a second or additional language has become an important concern in the field of assessment.

The need to better understand the relationship between language proficiency and content knowledge emerged initially in the context of bilingual education and the content-based instruction movement in the 1990s (Byrnes 2008). Since then, the need has only increased. As a result of immigration and globalization, a sizable proportion of students in schools and universities are learning content in a second or additional language. In the USA, for example, almost 10% of school-aged children are classified as English language learners (ELLs) (NCES 2015). Also, the workforce continues to become more global, and many workers carry out their profession in a second or additional language. In many parts of the world, English’s role as a lingua franca has meant that students often learn content in English in addition to their first language. The popularity of the content and language integrated learning (CLIL) movement in Europe, which involves the teaching and learning of content through a foreign language or lingua franca (typically English), is another example of a context in which language and content interact (Dalton-Puffer 2011). Finally, over the past couple of decades, there has been an increase in the number of English-medium universities (EMUs) and programs in places where English is a second or foreign language. English-medium education is most prevalent in Europe but is quickly expanding throughout the world (Wilkinson et al. 2006). Although important work on the relationship between language and content has been conducted in relation to CLIL and EMUs, the primary concerns in terms of assessment have been the language assessment policies and practices affecting the students and the faculty

in these programs. The focus has not yet shifted to the integration of language content in assessment (see Hofmannová et al. (2008) for emerging work on assessment that integrates language and content in a CLIL course). Wilkinson et al. (2006) assert that “the fact that education takes place through a language that is not the students’ mother tongue (and, in many cases, not that of the educators either) seems to have little influence on the assessment processes” (p. 30). They explain that “the typical approach would be to apply assessment processes that are virtually the same as would be applied in the mother tongue context” (pp. 29–30). Given that the focus of this chapter is on the relationship between language and content in construct definitions in assessment, the remainder of the chapter will focus on areas of research where this relationship has been explicitly explored.

Early Developments

Content-based instruction changed the landscape of language teaching by shifting the focus from communication in general to content as a context for language learning (Brinton et al. 1989). It is in the context of content-based instruction and bilingual education programs that concerns about the relationship between content and language began to be explicitly articulated (Byrnes 2008). As Short (1993) explains, in this context English learners needed to be involved in “regular curricula before they have fully mastered the English language” since “there simply is no time to delay academic instruction until these students have developed high levels of English language proficiency if they are to stay in school, succeed in their classes, and graduate with a high school diploma” (p. 628) – a claim still valid and relevant today for students around the world who are in school systems where they learn content in a second or additional language. Short strongly promotes the use of alternative assessments over standardized tests for assessing students in integrated language and content courses and programs, including the use of skill checklists and reading/writing inventories, anecdotal records and teacher observations, student self-evaluations, portfolios, performance-based tasks, essay writing, oral reports, and interviews. Even though she acknowledges “some overlap will occur between the language and content,” she argues that when it comes to assessment, “it is more advisable to focus on a single objective, be it content or language specific” (pp. 634–35).

Major Contributions

Major contributions to our understanding of the relationship between language proficiency and content knowledge in assessment emerged from the following areas of research: (1) language for specific purposes (LSP) testing and (2) content and language assessment of ELLs in schools.

Language for Specific Purpose Testing

The complicated relationship between content and language has long been acknowledged in the field of languages for specific purposes (LSP). Davies (2001), for example, argued that “LSP testing cannot be about testing for subject specific knowledge. It must be about testing the ability to manipulate language functions appropriately in a wide variety of ways” (p. 143). Douglas (2005), however, stated that the defining characteristic of LSP assessment is “a willingness, indeed a necessity, to include nonlinguistic elements in defining the construct to be measured” (p. 866). In fact, he argued that LSP testing “is defined by the nature of the construct to be measured, which includes both specific purpose language and background knowledge” (p. 866). One way in which background or content knowledge has been taken into account in LSP assessment is by incorporating “indigenous assessment criteria” (Jacoby and McNamara 1999), that is, assessment criteria derived from the target language use domain.

A recent example of a study that identifies the indigenous criteria that underlie professional judgments of communication in the context of the health professions is that of Elder et al. (2012). The rationale for their investigation, as for much of the work on LSP assessment, is that “if LSP tests are to act as proxies for the demands of communication faced by candidates entering the workforce, then the judgments of such professionals should not be ignored” (p. 409). In their study, they asked several health professionals to provide feedback on video recordings of trainee-patient interactions from the Occupational English Test, a specific-purpose English language test used in Australia for overseas-trained health professionals. Performances on this test are assessed using primarily linguistic criteria, including intelligibility, fluency, appropriateness of language, resources of grammar and expression, and overall communicative effectiveness.

They found that the health professionals in their study rarely mentioned language skills in their feedback about the performances they observed. The authors hypothesize that the health professionals’ lack of attention to language skills may be “because they give priority to clinical matters, because they feel that commenting on such features is beyond their competence, because they are blind to them (i.e., they lack the skills to make a linguistic diagnosis) or, more radically, because such features are irrelevant to what counts in clinical communication in their view” (p. 416). Elder et al. (2012) speculate that it may be that the candidates evaluated were already above a certain threshold of language proficiency that allowed the health professionals to focus on the clinical aspects of the performance. Uncovering the precise reasons for why the health professionals did not attend to language skills would be an important next step to better understand the role of content and language in this particular context.

Focusing on another LSP context, aviation English, Emery (2014) reflects on developments in the field in the last 30 years. He argues that the major change has been “the acceptance that it is neither possible nor desirable to separate language knowledge from subject matter knowledge” (p. 213). Nonetheless, he notes that “the extent and nature of the relationship between subject matter knowledge and

performance on language tests and the threat this represents to the validity of scores” continues to be a key issue in LSP testing. He explains, however, that in the case of aviation English where those assessed are trained and licensed professional pilots and air traffic controllers with high level of expertise in their field, “the question of whether it is possible or even desirable to separate subject matter knowledge from language knowledge is perhaps less relevant.” (Emery 2014, p. 210).

In fact, LSP testing in general often focuses on adults with high levels of expertise in a particular field. For this population, the challenge might simply be identifying the minimum threshold level of proficiency needed for communication. It may be that beyond that level of proficiency, language no longer plays an important role. The challenge for the field of LSP then would be identifying what that threshold is. Content and language assessment in schools, however, present different challenges in that students are developing both their language proficiency and their content knowledge at the same time.

Content and Language Assessment of ELLs in Schools

A greater focus on testing and accountability in many countries around the world has resulted in more assessments of students, including those learning in a second or additional language. In the USA, for example, No Child Left Behind (2001) required that all students including ELLs had to be assessed in the content areas of English language arts, mathematics, and science. The legislation also required that ELLs’ language proficiency had to be assessed annually. The need to assess all students in the content areas and the fact that a large proportion of students in schools are ELLs prompted discussions about the challenge of assessing ELLs’ content knowledge in English. Similarly, the need to annually assess ELLs’ language proficiency prompted discussions about the most appropriate and useful ways to do so. At the heart of these discussions was the content-language link.

Content-language link in content assessments. The main challenge in assessing ELLs in the content areas in English had been the score interpretation. Does the score on a content assessment represent the student’s content knowledge or does it represent their ability to read, understand, and respond to questions in English? Abedi (2004) argues that language is a source of construct-irrelevant variance when assessing ELLs in the content areas and that scores from these assessments are not meaningful indicators of students’ content knowledge. This perspective is supported by correlational studies that have found a relationship between the presence of complex linguistic features in test items and greater relative difficulty of the items for ELLs (e.g., Wolf and Leon 2009). Accommodations, modifications made to the assessment or the assessment administration, were introduced as a way to provide ELLs an opportunity to demonstrate their mastery of the content (Abedi et al. 2004). The assumption underlying accommodations is that language and content are separate constructs and that students will be able to demonstrate their content knowledge if their language ability does not get in the way.



RAHNAMA
P R E S S

@RAHNAMAPRESS
WWW.RAHNAMAPRESS.COM

