Encyclopedia of
Language and Education
Series Editor: Stephen May

SPRINGER
REFERENCE

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

Language Testing and Assessment

Third Edition

Springer

Encyclopedia of Language and Education

Series Editor
Stephen May
Faculty of Education and Social Work
The University of Auckland
Auckland, New Zealand

In this third, fully revised edition, the 10 volume *Encyclopedia of Language and Education* offers the newest developments, including an entirely new volume of research and scholarly content, essential to the field of language teaching and learning in the age of globalization. In the selection of topics and contributors, the Encyclopedia reflects the depth of disciplinary knowledge, breadth of interdisciplinary perspective, and diversity of sociogeographic experience in the language and education field. Throughout, there is an inclusion of contributions from non-English speaking and non-Western parts of the world, providing truly global coverage. Furthermore, the authors have sought to integrate these voices fully into the whole, rather than as special cases or international perspectives in separate sections. The Encyclopedia is a necessary reference set for every university and college library in the world that serves a faculty or school of education, as well as being highly relevant to the fields of applied and socio-linguistics. The publication of this work charts the further deepening and broadening of the field of language and education since the publication of the first edition of the Encyclopedia in 1997 and the second edition in 2008.

More information about this series at http://www.springer.com/series/15111

Elana Shohamy • Iair G. Or • Stephen May
Editors

Language Testing and Assessment

Third Edition

With 9 Figures and 3 Tables

Springer

Editors
Elana Shohamy
School of Education
Tel Aviv University
Tel Aviv, Israel

Iair G. Or
School of Education
Tel Aviv University
Tel Aviv, Israel

Stephen May
Faculty of Education and Social Work
The University of Auckland
Auckland, New Zealand

ISBN 978-3-319-02260-4 ISBN 978-3-319-02261-1 (eBook)
ISBN 978-3-319-02262-8 (print and electronic bundle)
DOI 10.1007/978-3-319-02261-1

Library of Congress Control Number: 2017934050

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

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, and 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 indispensible 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 Kramsch, 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 Iair 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 major 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 discoursal 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 changes are 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 tests and their political, social, educational, and ethical dimensions has made the field even 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 multiculturality. 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, "superdiverse" 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 skills 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 USA. He concludes with a set of principles regarding the need to limit the use of 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 certain 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 new 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

Acknowledgment

Thanks to the following for their key editorial support:
Consulting Editor: Nancy Hornberger
Editorial Assistant: Lincoln Dam

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 179
Glenn Fulcher

Methods of Test Validation 193
Xiaoming Xi and Yasuyo Sawaki

Qualitative Methods of Validation 211
Anne Lazaraton

Training in Language Assessment 225
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) 417
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

About the Editors

Elana Shohamy is a Professor of Multilingual Education at Tel Aviv University where she teaches and researches various sociolinguistic issues related to coexistence and rights in multilingual societies within the frameworks of critical language testing, language policy, migration, and linguistic landscape. Her current work in language testing focuses on multilingual assessment. Elana authored *The Power of Tests: A Critical Perspective on the Uses of Language Tests* (2001), *Language Policy: Hidden agendas and new approaches* (2006), and she edited various books on the above topics. She was the editor of the 2008 edition of the present volume on language testing and assessment of the *Encyclopedia of Language Education* (Springer). Elana served as the editor of the journal *Language Policy* (2007–2015) and the founder and current editor of the new journal *Linguistic Landscape*. Professor Shohamy was granted the 2010 ILTA Lifetime Achievement Award in Cambridge LTRC for her work on critical language testing.

Iair G. Or is a PhD candidate at the Tel Aviv University School of Education, Multilingual Education Program, researching Hebrew and Arabic language policy in Israel. His book *Creating a Style for a Generation: The Beliefs and Ideologies of Hebrew Language Planners* was published in Hebrew in 2016 and contains a historical analysis of the linguistic ideologies in the Hebrew language revival discourse in Ottoman and British-ruled Palestine. Or's published articles focused on critical comparative analysis of Arabic and Hebrew in the Israeli educational system and matriculation exams, as well as the implementation of international standards such as the CEFR. Or received the 2013 AAAL Graduate Student Award for his study on language ideologies in the planning of Modern Hebrew. He teaches courses on language testing and alternative assessment at the MA TESOL and Multilingual Education programs at Tel Aviv University and Kibbutzim College, Tel Aviv.

Stephen May is Professor of Education in Te Puna Wānanga (School of Māori and Indigenous Education) in the Faculty of Education and Social Work, University of Auckland, New Zealand. He is an international authority on language rights, language policy, bilingualism and bilingual education, and critical multicultural approaches to education and, to date, has published 15 books and over 90 articles and chapters in these areas. His key books include *The Multilingual Turn* (2014), *Language and Minority Rights* (2nd edition, 2012), and, with Christine Sleeter, *Critical Multiculturalism: Theory and Praxis* (2010). In addition to being Editor-in-Chief of the 3rd edition of the *Encyclopedia of Language and Education*, he is a Founding Editor of the interdisciplinary journal, *Ethnicities*, and was from 2005 to 2015 Associate Editor of *Language Policy*. He is also a Fellow of the American Educational Research Association (AERA). His homepage is http://www.education.auckland.ac.nz/uoa/stephen-may.

Advisory Board

External Reviewers

Language Testing and Assessment

Ofra Inbar-Lourie
Darla Deardorff
Pardee Lowe
Kate Menken
Durk Gorter
Andy Kirkpatrick
Anne Lazaraton
Aek Phakiti
Constant Leung
Meg Malone
Elana Spector-Cohen
Richard Hill
Sultan Turkan

Contributors

Jamal Abedi School of Education, University of California, Davis, CA, USA

Alison L. Bailey Human Development and Psychology Division, Department of Education, University of California, Los Angeles, CA, USA

Beverly Baker Official Languages and Bilingualism Institute, University of Ottawa, Ottawa, ON, Canada

Monica Barni University for Foreigners of Siena, Siena, Italy

Rachel L. Brooks Federal Bureau of Investigation, Washington, DC, USA

Carol A. Chapelle Applied Linguistics, Department of English, Iowa State University, Ames, IA, USA

Liying Cheng Faculty of Education, Queen's University, Kingston, ON, Canada

Alan Davies University of Edinburgh, Edinburgh, UK

Kristin J. Davin Foreign and Second Language Education, School of Education, Loyola University Chicago, Chicago, IL, USA

Catherine Elder School of Languages and Linguistics, University of Melbourne, Parkville, VIC, Australia

Janna Fox School of Linguistics and Language Studies, Carleton University, Ottawa, ON, Canada

Kellie Frost Language Testing Research Centre, School of Languages and Linguistics, University of Melbourne, Parkville, VIC, Australia

Glenn Fulcher English Department, School of Arts, University of Leicester, Leicester, UK

Danielle Guzman-Orth Educational Testing Service, Princeton, NJ, USA

Ofra Inbar-Lourie The School of Education, The Program for Multilingual Education, Tel Aviv University, Tel Aviv, Israel

Eunice Eunhee Jang Department of Applied Psychology and Human Development, Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, Canada

Jennifer Jenkins Department of Modern Languages, University of Southampton, Southampton, UK

James P. Lantolf Language Acquisition and Applied Linguistics, Department of Applied Linguistics, Center for Language Acquisition, CALPER, The Pennsylvania State University, University Park, PA, USA

Anne Lazaraton Department of Writing Studies, University of Minnesota, Minneapolis, MN, USA

Constant Leung Centre for Language, Discourse and Communication, School of Education, Communication and Society, King's College London, London, UK

Jo Lewkowicz University Council for the Certification of Language Proficiency, University of Warsaw, Warszawa, Poland

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

Alexis A. Lopez Educational Testing Service, Princeton, NJ, USA

Margaret E. Malone Assessment and Evaluation Language Resource Center, Department of Linguistics, Georgetown University, Washington, DC, USA

Kate Menken Department of Linguistics, Queens College, City University of New York, Flushing, NY, USA

Research Institute for the Study of Language in an Urban Society, Graduate Center, City University of New York, New York, NY, USA

Matthew E. Poehner World Languages Education and Applied Linguistics, Department of Curriculum and Instruction, The Pennsylvania State University, University Park, PA, USA

Luis E. Poza School of Education and Human Development, University of Colorado Denver, Denver, CO, USA

James Enos Purpura Teachers College, Columbia University, New York, NY, USA

Luisa Salvati University for Foreigners of Siena, Siena, Italy

Yasuyo Sawaki Faculty of Education and Integrated Arts and Sciences, Waseda University, Tokyo, Japan

Angela Scarino Research Centre for Languages and Cultures, School of Communication, International Studies and Languages, Adelaide, SA, Australia

Elana Shohamy School of Education, Tel Aviv University, Tel Aviv, Israel

Megan Smith Department of English, Mississippi State University, East Lansing, MS, USA

Bernard Spolsky Bar-Ilan University, Ramat Gan, Israel

Charles W. Stansfield Language Learning and Testing Foundation, Rockville, MD, USA

Dina Tsagari Department of English Studies, University of Cyprus, Nicosia, Cyprus

Sultan Turkan Educational Testing Service, Princeton, NJ, USA

Guadalupe Valdés Race, Inequality, Language and Education (RILE) Program, Graduate School of Education, Stanford University, Stanford, CA, USA

Erik Voss NU Global, Northeastern University, Boston, MA, USA

Gillian Wigglesworth Research Unit for Indigenous Language, ARC Centre of Excellence for the Dynamics of Language, Faculty of Arts, University of Melbourne, Parkville, VIC, Australia

Xiaoming Xi New Product Development, Educational Testing Service, Princeton, NJ, USA

Part I

Assessing Language Domains

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

E. Shohamy et al. (eds.), *Language Testing and Assessment*, Encyclopedia of Language and Education, DOI 10.1007/978-3-319-02261-1_33

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 and 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 test 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.

However, research on the effectiveness of accommodations meant to reduce the linguistic load of test items has yielded mixed results, raising questions about this assumption (Kieffer et al. 2009, 2012). Outcomes of this research have led to a consensus on the need to better understand the language-content link. At minimum, it is important to distinguish "between language abilities central to the academic skills being measured and language demands of the test that are not relevant to the skills and abilities being measured" (Kieffer et al. 2012, p. 3). Avenia-Tapper and Llosa (2015) propose an approach for making distinctions between construct-relevant and construct-irrelevant language in content assessments. Drawing from systemic functional linguistics, they argue that certain complex linguistic features are a component of content area mastery, and thus, complex linguistic features cannot be considered construct-irrelevant on the basis of their complexity alone. Instead, the strong presence or absence of the linguistic features in the domain to which the test should generalize (e.g., grade-level science talk and text) is a better criterion for judging the relevance of a given linguistic feature. This approach would prevent assessment developers from eliminating complex structures that may be critical to the content area, thus guarding against the possibility of creating accommodated tests that suffer from construct underrepresentation, which could in turn cause negative washback for ELLs.

Content-language link in English language proficiency (ELP) assessments. Research on English language proficiency tests developed prior to NCLB uncovered that the language assessed by these tests did not align with the types of academic language that students needed to succeed in school (e.g., Stevens et al. 2000). Work was carried out to define and operationalize the construct of academic language proficiency by investigating empirically the kinds of English required of K–12 ELLs (Bailey and Butler 2003). Various categorizations of academic language emerged, describing it in terms of its lexical, grammatical, and textual characteristics (Bailey 2007).

An important shift in thinking about academic English proficiency was reflected in the ELP standards that emerged in 2004, which differed markedly in their conceptualization of English proficiency from most existing ELP standards. The existing ELP standards focused on language as communication and tended to be closely aligned to English language arts standards. The ELP standards, developed by the WIDA consortium (2004, 2007) and then augmented and adopted by TESOL (2006), were designed to link ELP to social and instructional language and to four content areas – language arts, mathematics, science, and social studies. Research on the ACCESS for ELLs, the ELP assessment designed to measure students' mastery of the WIDA standards, revealed that even though the assessment taps primarily into a language construct, content is assessed to some extent as well, especially at the higher levels of English proficiency. Romhild et al. (2011) identified "domain-general" and "domain-specific" linguistic knowledge factors underlying the structure of various forms (by grade and level of language proficiency) of this ELP assessment. Domain-general linguistic knowledge referred to academic language common to various content areas, whereas domain-specific knowledge referred to academic language specific to a particular content area. They found that the domain-

general factor was stronger in most forms of the test, but in forms assessing higher levels of English proficiency, the domain-specific factor was stronger than the domain-general factor. In other words, in assessments focused on students' mastery of ELP standards that link language proficiency to the content areas, it became difficult to disentangle language proficiency from content knowledge at higher levels of language proficiency.

Work in Progress

Work in progress in the area of K–12 assessment in the USA has the potential to inform and transform the way the relationship between language proficiency and content knowledge is envisioned and how these constructs will be assessed in the future. A new wave of standards has emerged through the Common Core State Standards (CCSS) for English language arts and literacy in history, social studies, science, and technical subjects, the CCSS in mathematics (Common Core State Standards Initiative 2010a, b), and the Next Generation Science Standards (NGSS Lead States 2013). A major feature of these new standards is an emphasis on literacy and *practices* that are language and discourse rich. For example, "engage in argument from evidence" is one of the NGSS practices, "comprehend as well as critique, value evidence" are included in the CCSS for English language arts, and "construct viable arguments and critique the reasoning of others" is in the CCSS in mathematics (Stage et al. 2013).

These standards represent a major shift in the way content is defined, taught, and assessed. For the past decade, ELP standards have moved toward the content areas, whereas now standards in the content areas are moving toward language. Responding to the demands of these new standards for all students and using these demands as opportunities to help ELLs will require new ways of thinking about the relationship between language and content learning (Valdés et al. 2014). The *Understanding Language* Initiative at Stanford University (http://ell.stanford.edu) has led the effort to support ELLs in meeting new content standards, adopting a view of language as action that focuses on the essential role of language in learning academic content. For example, as part of the *Understanding Language* Initiative, in the area of science education for ELLs, Lee et al. (2013) suggest "(a) a shift away from both content-based language instruction and the sheltered model to a focus on language-in-use environments and (b) a shift away from 'teaching' discrete language skills to a focus on supporting language development by providing appropriate contexts and experiences" (p. 228). They introduce a conceptual framework that illustrates how the science and engineering practices in the NGSS can be unpacked into the types of language and discourse needed to instantiate these practices.

Two consortia are developing ELP assessments that are aligned to the new content standards. WIDA revised its standards and its assessment, ACCESS for ELLs. The revised standards still link language proficiency to social and instructional language and the four content areas, but are more explicit about how academic language is conceptualized by outlining specific features at the word/phrase,

sentence, and discourse level. At the word/phrase level, the focus is on vocabulary usage; at the sentence level, the focus is on language forms and conventions; and at the discourse level, the focus is on linguistic complexity. The second consortium, the English Language Proficiency Assessment for the twenty-first century (ELPA21) consortium, has developed ELP standards as a foundation to their assessment system informed by the work of the *Understanding Language Initiative*. ELPA21 specified ten standards that focus on form (e.g., vocabulary, grammar, and discourse specific to particular content areas) and function (e.g., what students do with language to accomplish content-specific tasks).

Problems and Difficulties

The lack of empirical research about the development of and relationship between content knowledge and language proficiency remains a major challenge. Byrnes (2008) explains that "because content knowledge in an L2 learning environment is even more a developmental matter than is the case for native language instruction, content assessment would benefit from principles that identify how content and language abilities develop simultaneously in language learning" (p. 45). In the context of K-12 assessment in schools specifically, there is a lack of research on the relationship between (academic) language development and content instruction for all students, not just ELLs (Frantz et al. 2014). This lack of empirical research and the fact that both language proficiency and content knowledge develop across grades makes it particularly difficult to establish boundaries between these constructs. These boundaries are important as long as there is a need or mandate to assess language proficiency and content knowledge separately as is the case in the USA and in many other countries. Another reason why it might be important to locate these boundaries is to be able to use assessment information diagnostically. It may be helpful for educators to be able to identify sources of students' difficulty in accomplishing a task, whether it be language, content, or both.

Future Directions

New task types and advances in technology may allow us to better understand the content-language link and develop assessments that assess content and language in an integrated way and at the same time allow for some separation of the two constructs. Integrated tasks, tasks that assess more than one language skill, have been developed in the past several years in response to increased awareness of the complexity of language use and the importance of context. The TOEFL iBT, for example, includes integrated tasks that require students to read a passage, listen to a lecture, and respond in writing. Integrated tasks are believed to be more representative of actual language use and thus allow for score-based interpretations that can be generalized to a particular target language use domain. A similar rationale could be

applied to justify the development of integrated tasks of language proficiency and content knowledge.

Given the focus on language and literacy skills in the content areas in the new standards, new content assessments will need to embrace these broader definitions of content and engage students' rich language use. Thus, a separate assessment of ELP may not be needed; it may be possible to assess language proficiency and content knowledge within the same assessment (Bailey and Wolf 2012). One technology-based innovation that would lend itself to integrated assessments of language and content are scenario-based assessments. These types of assessments are specifically designed to assess learners' integrated skills in a purposeful, interactive, and strategic manner. Scenario-based assessments have been used primarily to assess reading skills (Sabatini et al. 2014), but their use for ELP assessment is already being explored. In the content area of science, simulation-based assessments have been developed for both high-stakes summative assessment and classroom formative assessment (e.g., Quellmalz et al. 2012). Simulation-based assessments allow students to demonstrate their science knowledge as well as their ability to engage in scientific practices (e.g., predicting, observing, explaining findings, arguing from evidence). It may be possible to add a language dimension to these simulations so that language skills, which are already elicited as part of the assessment of science practices, are assessed alongside science content.

Finally, another innovation in assessment that would make integrated assessments of content and language particularly useful for instructional and diagnostic purposes is the use of scaffolds embedded in technology-enhanced assessments. Wolf and Lopez (2014) have examined the impact of including scaffolds in a scenario-based assessment of young ELLs' language proficiency. Their assessment includes speaking tasks with scaffolding questions: Students first retell a story independently, then answer scaffolding questions, and then retell the story for a second time. They found that students were more successful in retelling the story after responding to the scaffolding questions and that low-performing students on the task were at least able to complete the scaffolding questions. They concluded that "the incorporation of scaffolding into assessment has the potential to improve the measurement of EL students' language proficiency and also provide useful information for teachers' instruction." Both content and language scaffolds could be incorporated into technology-enhanced and scenario- or simulation-based assessments. In fact, simulation-based assessments already have the capability to provide scaffolds and immediate feedback and coaching related to the science knowledge and inquiry practices being assessed by the simulations (Quellmalz et al. 2012). These assessments have also experimented with accommodations for ELLs and student with disabilities, including audio recordings of text, screen magnification, and segmentation to support reentry at the beginning of a task to allow for extended time (Quellmalz et al. 2012). Much more refined language scaffolds could be added to these simulations to allow ELLs of different levels of proficiency to engage with the tasks and demonstrate both their content knowledge and their language proficiency. Scaffolds could be informed by current work on learning progressions in the content areas (see NGGS Lead States 2013 as an example) and in specific areas of language

development (Bailey and Heritage 2011–15). These types of innovative, technology-enhanced, simulation-based, scaffolded assessments could be used both to assess and promote learning and also as a means to investigate the developmental nature of content and language learning for ELLs.

Cross-References

- ▶ Assessing Meaning
- ▶ Dynamic Assessment
- ▶ Language Assessment in Higher Education
- ▶ Language Assessment Literacy

Related Articles in the Encyclopedia of Language and Education,

Masaki Kobayashi, Sandra Zappa-Hollman, Patricia Duff: Academic Discourse Socialization. In Volume: Language Socialization

Mary R. Lea: Academic Literacies in Theory and Practice. In Volume: Literacies and Language Education

Tarja Nikula: CLIL: A European Approach to Bilingual Education. In Volume: Second and Foreign Language Education

Fredricka L. Stoller, Shannon Fitzsimmons-Doolan: Content-Based Instruction. In Volume: Second and Foreign Language Education

References

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4–14.

Abedi, J., Hofstetter, C., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1–28.

Avenia-Tapper, B., & Llosa, L. (2015). Construct relevant or irrelevant? The role of linguistic complexity in the assessment of English language learners' science knowledge. *Educational Assessment, 20*, 95–111.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bailey, A. L. (2007). *The language demands of school: Putting academic English to the test*. New Haven: CT Yale University Press.

Bailey, A. L., & Butler, F. A. (2003). *An evidentiary framework for operationalizing academic language for broad application to K-12 education: A design document* (CSE Technical Report No. 611). Los Angeles: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

Bailey, A. L., & Heritage, M. (2011–15). Dynamic Language Learning Progressions Project. www.dllp.org

Bailey, A. L., & Wolf, M. K. (2012). The challenge of assessing language proficiency *aligned* to the Common Core State Standards and some possible solutions. Retrieved from http://ell.stanford.edu/papers/policy

Brinton, D., Snow, M. A., & Wesche, M. B. (1989). *Content-based second language instruction*. New York: Newbury House.

Byrnes, H. (2008). Assessing content and language. In E. Shohamy and N. H. Hornberger (Eds.), *Encyclopedia of language and education, 2nd edition, volume 7: Language testing and assessment* (pp. 37–52). New York: Springer Science+Business Media.

Common Core State Standards Initiative. (2010a). Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects. Retrieved from http://www.corestandards.org/ELA-Literacy

Common Core State Standards Initiative. (2010b). Common Core State Standards for mathematics. Retrieved from http://www.corestandards.org/Math

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics, 31*, 182–204.

Davies, A. (2001). The logic of testing languages for specific purposes. *Language Testing, 18*(2), 133–147.

Douglas, D. (2005). Testing languages for specific purposes. In E. Hinkel (Ed.), *The handbook of research in second language teaching and learning* (pp. 857–868). Mahwah: Erlbaum.

Elder, C., Pill, J., Wookward-Kron, R., McNamara, T., Manias, E., Webb, G., & McColl, G. (2012). Health professionals' views of communication: Implications for assessing performance on a health-specific English language test. *Tesol Quarterly, 46*(2), 409–419.

Emery, H. J. (2014). Developments in LSP testing 30 years on? The case of aviation English. *Language Assessment Quarterly, 11*(2), 198–215.

Frantz, R. S., Bailey, A. L., Starr, L., & Perea, L. (2014). Measuring academic language proficiency in school-age English language proficiency assessments under new college and career readiness standards in the United States. *Language Assessment Quarterly, 11*, 432–457.

Hofmannová, M., Novotná, J., & Pípalová, R. (2008). Assessment approaches to teaching mathematics in English as a foreign language. *International CLIL Research Journal, 1*(1), 20–35.

Jacoby, S., & McNamara, T. (1999). Locating competence. *English for Specific Purposes, 18*, 213–241.

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta analysis on effectiveness and validity. *Review of Educational Research, 79*(3), 1168–1201.

Kieffer, M. J., Rivera, M., & Francis, D. J. (2012). *Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments. 2012 update*. Portsmouth: RMC Research Corporation, Center.

Lead States, N. G. S. S. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners in relation to next generation science standards and with implications for common core state standards for English language arts and mathematics. *Educational Researcher, 42*(4), 223–233.

National Center for Education Statistics. (2015). The Condition of Education 2015 (NCES 2015–144). Retrieved from https://nces.ed.gov/programs/coe/indicator_cgf.asp.

No Child Left Behind Act. (2001). Pub. L. No. 107–110, § 115, Stat. 1425(2002).

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching, 49*(3), 363–393.

Romhild, A., Kenyon, D., & McGregor, D. (2011). Exploring domain-general and domain-specific linguistic knowledge in the assessment of academic English language proficiency. *Language Assessment Quarterly, 8*(3), 213–228. doi:10.1080/15434303.2011.558146.

Sabatini, J. P., O'Reilly, T., Halderman, L. K., & Bruce, K. (2014). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disabilities Research & Practice, 29*(1), 36–43.

Short, D. J. (1993). Assessing integrated language and content instruction. *TESOL Quarterly, 27*, 627–656.

Stage, E. K., Asturias, H., Cheuk, T., Daro, P. A., & Hampton, S. B. (2013). Opportunities and challenges in next generation standards. *Science, 340*, 276–277.

Stevens, R., Butler, F. A., & Castellon-Wellington, M. (2000). *Academic language and content assessment: Measuring the progress of ELLs*. Final Deliverable to OERI, Contract No. R305B60002. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Teachers of English to Speakers of Other Languages. (2006). *PreK-12 english language proficiency standards*. Alexandria, Virginia: TESOL.

Valdés, G., Kibler, A., & Walqui, A. (2014). *Changes in the expertise of ESL professionals: Knowledge and action in an era of new standards*. Alexandria: TESOL International Association.

Wilkinson, R., Zegers, V., & van Leeuwen, C. (2006). *Bridging the assessment gap in English-Medium higher education*. Maastricht University Language Centre.

Wolf, M., & Leon, S. (2009). An investigation of the language demands in content assessments for English language learners. *Educational Assessment, 14*(3), 139–159.

Wolf, M., & Lopez, A. (2014). *The use of scaffolding strategies in the assessment of English Learners*. Paper presented at the Roundtable on Learning-Oriented Assessment in Language Classrooms and Large-Scale Assessment Contexts in New York.

Culture and Language Assessment

Angela Scarino

Abstract

This first entry on culture and language assessment is written at a time of much reconsideration of the major constructs in language/s learning and language assessment. This is in response at least partly to the increasingly complex reality of multilinguality and multiculturality in our contemporary world. Culture is one of these constructs and is considered in its interrelationship with language and learning. It is because of this reconsideration that the discussion in this chapter is focused on scoping the conceptual landscape and signaling emerging rather than established lines of research. The discussion encompasses (a) the assessment of culture in the learning of languages, including recent interest in assessing intercultural practices and capabilities, and (b) the role of culture (and language), or its influence, on the assessment of learning where multiple languages are in play. The discussion considers the place of culture in conceptualizing the communicative competence and understandings of the role of culture in all learning. Developments related to the assessment of intercultural practices and capabilities in foreign language learning are described, as well as multilingual (and multicultural) assessment approaches. The assessment of capabilities beyond the linguistic poses major challenges to traditional conceptualizations and elicitation and judgment practices of assessment. This is because what is being assessed is the linguistic and cultural situatedness of students of language/s as they communicate and learn across linguistic and cultural systems. This challenges the traditional assessment paradigm and also raises important ethical issues. This conceptual and practical stretch can only extend thinking about educational assessment.

A. Scarino (✉)
Research Centre for Languages and Cultures, School of Communication, International Studies and Languages, Adelaide, SA, Australia
e-mail: angela.scarino@unisa.edu.au

E. Shohamy et al. (eds.), *Language Testing and Assessment*, Encyclopedia of Language and Education, DOI 10.1007/978-3-319-02261-1_3

Keywords

Assessing culture • Culture in assessment • Multilingual assessment • Intercultural capabilities • Translanguaging • Hermeneutic approaches

Contents

Introduction 16
Early Developments 17
Language and Culture 17
Language, Culture, and Learning 19
Major Contributions 20
Ongoing Conceptualization of Communicative Competence Toward "Intercultural Competence" 20
Assessing Intercultural Competence 21
Multilingual Assessment Approaches 23
Work in Progress 24
Problems and Difficulties 26
Conceptualizing Culture and Language Assessment 26
Elicitation 26
Judging 27
Future Directions 28
Cross-References 28
Related Articles in the Encyclopedia of Language and Education 28
References 29

Introduction

It is timely to consider culture and language assessment, as culture is a dimension that has been undergoing major reconsideration in language/s learning in the past decade (e.g., see Byrnes 2010), and yet it is underrepresented in the language assessment literature.

The discussion in this chapter will consider mainly the assessment of culture/s in the learning of language/s, including the recent interest in assessing intercultural practices and capabilities in language/s learning. This refers to how "cultural knowing" or "cultural/intercultural understanding" is assessed in the context of learning language/s. The discussion will also consider, to a lesser extent, the role of culture, or its influence, on the assessment of learning in environments where multiple languages are in play and where students are or are becoming multilingual. This aspect highlights that the process of assessing learning (of language/s or other disciplines) is itself both a cultural (and linguistic) act and that culture/s (and language/s) come into play in learning and in the assessment of learning. This is because students are linguistically and culturally situated in the linguistic and cultural systems of their primary socialization. In developing their learning and in assessment, they draw upon their own dynamic histories of experiences of knowing, being, and communicating and their own frameworks of values and dispositions. In discussing both aspects, the focus will remain specifically on education and educational assessment.

The discussion takes as a starting point the move in language/s teaching and learning, away from a monolingual and national paradigm (with the one language equals one culture equation) toward a multilingual paradigm. [For a detailed discussion, see the guest-edited volume of The Modern Language Journal by Claire Kramsch (MLJ, 98, 1, 2014), "Teaching foreign languages in the era of globalisation."] It is this move that gives greater prominence to the interplay of multiple languages and with these multiple cultures, in all learning and therefore in assessment.

In a recent 25-year review of culture in the learning of foreign languages, Byram (2014), one of the most prolific writers on the role of culture in language teaching, learning, and assessment, observed that "the question of assessment remains insufficiently developed"(p. 209). Atkinson (1999) reflected on how little direct attention is given to the notion of culture in TESOL, even though "ESL teachers face it in everything they do" (p. 625). Block (2003), discussing the social turn in second language acquisition (SLA), raised questions about "a cultural turn" for SLA research. He specifically noted the difficulty involved in conceptualizing the relationship between language and culture, but also the promising work in pragmatics and in learner identity as research areas that take culture into account. Although Shohamy (2011) did not specifically address culture, she drew attention to an important dimension of the discussion when arguing for assessing "multilingual competencies" in an assessment field that continues to view language as a monolingual, homogenous, and often still native-like construct (p. 419). I add that the monolingual bias that Shohamy described in language assessment extends to it also being a *monocultural* bias.

These reflections signal some of the efforts to reconsider and expand the constructs in language teaching, learning, and assessment "beyond lingualism" (Block 2014) to include dimensions such as subjectivity and identity.

Early Developments

Culture comes into play in the diverse contexts of language learning and assessment, both as a dimension of the *substance* of learning (e.g., in the learning of foreign languages) and as the *medium* for learning language/s and other areas of learning (e.g., in the learning of ESL/EAL). In considering culture as substance, it is necessary to consider the relationship between language and culture. In considering culture as medium, it is necessary to consider the relationship between language and culture and learning.

Language and Culture

The integral relationship between language/s and culture/s has long been considered from diverse disciplinary perspectives, including linguistics, anthropology, sociolinguistics, and applied linguistics (Whorf 1940/1956; Sapir 1962; Geertz 2000; Gumperz and Levinson 1996; Kramsch 2004). In the diverse contexts of language/s

learning, this interrelationship is understood and foregrounded in different ways. In foreign language teaching and learning, culture has been understood traditionally as factual knowledge or as a form of "content" of language learning, with literature and other aesthetic forms as rich expressions of particular culture/s. In this sense culture is understood as observable products or artifacts, associated with a particular social group. It has also been understood as ways of life, behaviors, and actions of a social group where the language/s is used. Both of these understandings present a static view of culture that removes variability and personal agency within the national group. A more recent perspective is an understanding of culture as social norms and practices, created through the use of language (see Byrnes 2010). Such practices, however, are removed from the cultural identity of the learner as a participant in language learning. In ESL/EAL, where the major goal is to prepare students for learning in English across diverse disciplines, the interrelationship between language and culture has been backgrounded in order to focus on subject matter learning across the curriculum.

A useful starting point for a consideration of culture and language assessment is how it has been represented in the construct of "communicative competence." This is because it is the conceptualization of the construct that guides elicitation, judging, and validation in the assessment cycle (Scarino 2010). In the conceptualization of "linguistic competence," where the focus was on the linguistic system itself, there was an absence of any attention to culture or to language users as participants in the linguistic and cultural system. Canale and Swain's (1980) framework comprised grammatical competence (vocabulary and rules of grammar), sociolinguistic competence (conventions of use), discourse competence (cohesion and coherence of texts), and strategic competence (compensating for limited resources in using language). This modeling highlighted the social, interactive nature of language use and the crucial role of context. The sociolinguistic interest here was with how the social context affects choices within the linguistic system. Halliday's theoretical work is instructive in this regard.

Halliday (1999) used the theoretical constructs "context of situation"' and "context of culture" to explain what is entailed in an exchange of meanings in communication. In Halliday's terms, these two constructs do not refer to "culture" in the sense of lifestyles, beliefs, and value systems of a language community (e.g., as in traditional foreign language learning) but rather as a system of meanings. He makes clear that the two constructs are not two different things, but rather that they are the same thing seen from two different depths of observation. The "situation" provides the context for particular instances of language use, and, as such, it is an instance of the larger system, which is referred to as "culture." For Halliday, culture is in the very grammar that participants use in exchange.

Bachman and Palmer (1996, 2010) built on the Canale and Swain model by identifying "knowledge" in the mind of the user, which can be drawn upon in communication. They identified (a) organizational knowledge, that is, grammatical knowledge and textual knowledge, and (b) pragmatic knowledge, that is, functional and sociolinguistic knowledge. Pragmatic knowledge is understood as objective

knowledge that is necessary for selecting language appropriately for use in particular social situations. As such, it represented a static view of the context of situation and of participants in that context. Although this is recognized as the most developed model of "communicative ability" for the purposes of assessment, it has been criticized because of its individualistic view of social interaction (McNamara and Roever 2006) and because context is not sufficiently taken into account (Chalhoub–Deville 2003; see also Bachman 2007). In the extensive discussion about context in defining the construct of communicative competence in language assessment, the context has been understood essentially as the context of situation, with little explicit attention to the context of culture.

The applied linguist who has most extensively theorized culture in (foreign) language learning is Claire Kramsch. In her 1986 critique of the proficiency movement as an oversimplification of human interaction, Kramsch extended the construct from communicative to "interactional competence." She highlighted at the same time that this interaction takes place within "a cross-cultural framework" (p. 367) and that successful interaction necessitates the construction of a shared internal context or "sphere of intersubjectivity" (p. 367). This understanding of culture foreshadowed her extensive discussion of context and culture in language teaching (Kramsch 1993) and her subsequent theorization of culture as "symbolic competence" (Kramsch 2006), which I consider below (see section "Major Contributions").

Language, Culture, and Learning

Language and culture are integral to learning. Halliday (1993) highlighted learning itself as a process of meaning-making when he wrote:

> When children learn a language, they are not simply engaging in one kind of learning among many; rather, they are learning the foundation of learning itself. The distinctive characteristic of human learning is that it is a process of meaning making – a semiotic process. (p. 94)

It is through language, in the context of situation and the context of culture, that students and teachers, in their diversity, interact to exchange knowledge, ideas, explanations, and elaborations and make sense of and exchange meaning in learning. In the learning interaction, this meaning is mediated through the lenses of the language/s and culture/s of participants' primary socialization.

All learning, therefore, is essentially a linguistic and cultural activity. It is formed through individual learners' prior knowledge, histories, and linguistic and cultural situatedness. It is the learner's situatedness and the cultural framing of learning that shapes the interpretation and exchange of meanings in learning and, by extension, in the assessment of learning. This understanding is in line with cultural views of learning in education. Gutierrez and Rogoff (2003) described learning as emerging from participating in practices, based on students' linguistic and cultural–historical repertoires. Lee (2008) also discussed "the centrality of culture to . . . learning and

development" (p. 267). This understanding of the relationship between language, culture, and learning is related to the sociocultural family of theories of language learning, in which the role of culture at times remains implicit. This understanding of learning as a linguistic, social, and cultural act of meaning-making becomes important in assessment. Shohamy (2011) expressed concern with the differential performance of immigrant students, depending on whether they are assessed in the language of their primary socialization or in the language of education in their new locality. The meanings that students make and represent in learning and assessment necessarily originate in the linguistic, cultural, experiential, and historical knowledge context to which they belong. It is this relationship that underlies Shohamy's argument for multilingual assessment (see section "Future Directions" below).

Major Contributions

Major contributions to the consideration of culture and language assessment have been advanced in relation to ongoing conceptualizations of the construct of communicative competence, including toward "intercultural competence," the assessment of intercultural practices and capabilities, and multilingual approaches to assessment.

Ongoing Conceptualization of Communicative Competence Toward "Intercultural Competence"

In more recent work, Kramsch has expanded further the constructs of communicative competence and interactional competence to what she has termed "symbolic competence" (Kramsch 2006). In her conceptualization, knowledge of and engagement with the systems of culture associated with language provide the basis for understanding the ways in which users of the language establish shared meanings, how they communicate shared ideas and values, and how they understand the world. Language constitutes and reflects the social and cultural reality that is called context. Symbolic competence foregrounds meaning-making not only as an informational exchange but as a process of exchange of cultural meaning, including its interpretive and discursive symbolic dimensions. It entails using language to negotiate and exchange meanings in context, both reciprocally with others and in individual reflection on the nature of the exchanges. Context is not fixed or given but created in interaction through the intentions, assumptions, and expectations of participants. Kramsch foregrounded not only such exchange *within* a language but also *across* languages and cultures in multilingual and multicultural contexts, and it is in this way that she elaborated foreign language learning as an intercultural endeavor that develops "intercultural competence."

Assessing Intercultural Competence

Perhaps because her conceptualization of culture and the intercultural in language/s learning is the most elaborated and complex, Kramsch (2009) questioned whether or not it can be assessed. She stated:

> [S]ymbolic competence based on discourse would be less a collection of. . . stable knowledges and more a savviness i.e., a combination of knowledge, experience and judgment. . . Trying to test symbolic competence with the structuralist tools employed by schools. . . is bound to miss the mark. Instead, symbolic competence should be seen as the educational horizon against which to measure all learners' achievements. (p. 118)

This may well be the case within traditional testing paradigms, but it has been suggested that possibilities may be available within alternative assessment paradigms (Scarino 2010) and assessment purposes that are educational.

In considering assessment in the context of intercultural language learning, a major distinction needs to be drawn between the consideration of "intercultural understanding" in general education, where language is not foregrounded (Bennett 1986), and in language/s education, where language use and language learning are the focus.

The extensive efforts to model intercultural competence began with Byram and Zarate (1994) and Byram (1997), working under the auspices of the Council of Europe. Their conceptualization was based on a set of knowledge, skills, and dispositions called *savoirs*: *savoir apprendre*, *savoir comprendre/faire*, *savoir être*, and *savoir s'engager*. In line with the council's orientation, it was focused on an objectives-setting approach, which was analytic rather than holistic, and on defining levels of intercultural competence. Although these *savoirs* captured broad educational dimensions such as *savoir être* (knowing how to be) and *savoir s'engager* (knowing how to engage politically), the original modeling did not sufficiently foreground communication. Byram (1997) subsequently modeled "intercultural communicative competence," incorporating the set of dimensions of the model of Canale and Swain (1981), discussed above, with the set of *savoirs* that defined intercultural competence. As with all modeling, however, the relationship among these sets of dimensions was not explained. Risager (2007) included further dimensions, which she described as "linguacultural competence," resources, and transnational cooperation, thereby highlighting the multilingual (and multicultural) nature of communication. Sercu (2004) considered the inclusion of a "metacognitive dimension" that focuses on students monitoring their learning. Although this is a valuable dimension, Sercu did not specify that the reflective work should be focused on exploring the linguistic and cultural situatedness of participants involved in communication and learning to communicate interculturally, and how it is this situatedness that shapes the interpretation, creation, and exchange of meaning. The consideration of the intricate entailments of this intercultural capability was extended by Steffensen et al. (2014) to include timescales and identity dynamics. The focus

specifically on identity formation was also taken up by Houghton (2013), with what she refers to as *savoir se transformer*.

In conceptualizing intercultural competence (or more precisely, "interlinguistic and intercultural practices and capabilities") for the purposes of assessment, Liddicoat and Scarino (2013) highlighted the need to capture:

- Observation, description, analysis, and interpretation of phenomena shared when communicating and interacting
- Active engagement with the interpretation of self (intraculturality) and "other" (interculturality) in diverse contexts of exchange
- Understanding the ways in which language and culture come into play in interpreting, creating, and exchanging meaning
- The recognition and integration into communication of an understanding of self (and others) as already situated in one's own language and culture when communicating with others
- Understanding that interpretation can occur only through the evolving frame of reference developed by each individual (pp. 130–131)

Assessment in this formulation, therefore while remaining focused on language and culture, encompasses more than language. It is at once experiential, analytic, and reflective. For Liddicoat and Scarino (2013), it includes (a) language use to communicate meanings in the context of complex linguistic and cultural diversity, with a consideration of both personal and interpersonal subjectivities, (b) analyses of what is at play in communication that is situated within particular social and realities and how language and culture come into play in the practice of meaning-making, and (c) reciprocal reflection and reflexivity in relation to self as intercultural communicator and learner.

In addition to extensive work on conceptualizing the assessment of intercultural practices and capabilities, practical work has been and continues to be undertaken to develop ways of eliciting these practices and capabilities (e.g., see, Byram 1997; Deardorff 2009; Lussier et al. 2007). Sercu (2004) attempted to develop a typology of assessment tasks including five task types: cognitive, cognitive-attitudinal, exploration, production of materials, and enactment tasks. This framework, however, does not address precisely these capture intercultural practices and capabilities.

As indicated, it is the alternative qualitative assessment paradigm, particularly within a hermeneutic perspective (Moss 2008) and inquiry approaches (Delandshere 2002), which offers the most fruitful basis for considering the assessment of these practices and capabilities in language/s learning. Liddicoat and Scarino (2013, chapter 8) discussed and illustrated ways of eliciting the meanings that learners make or accord to phenomena and experiences of language learning, and their analyses and reflections on meaning-making. The learner is positioned as performer and analyzer, as well as being reflective. An issue that remains to be considered with respect to elicitation is the complex one of integrating the performative, analytic, and reflective facets.

The area of judging is possibly the most complex of all, not only because educators hesitate to assess learner subjectivity and the realm of values and dispositions but also because of the difficulty of bringing together, in some way, the diverse facets of intercultural practices and capabilities. Although a framework for setting criteria for judging performance has been proposed (see Liddicoat and Scarino 2013, pp. 138–139), the extent to which criteria can be pre-specified or else should emerge from the specific context of the exchange still needs to be addressed.

Finally, there is not yet in the field a frame or frames of reference for making judgments of such practices and capabilities. The Council of Europe has sought to develop a scale to address this absence but efforts to date have not succeeded. This is not surprising given the complexity that this would entail. Although making judgments remains an area of uncertainty for assessors, it is not likely to be resolved by a generalizing scale.

Multilingual Assessment Approaches

"Multilingual assessment" is a practice proposed by Shohamy (2011) that would take into account all the languages in the multilingual speaker's repertoire as well as "multilingual functioning" (Shohamy 2011, p. 418). Given the interrelationship between language/s and culture/s discussed above, this multilingual functioning also implies *multicultural* functioning. It is useful to distinguish at least two senses of multilingual assessment. The first is multilingual in the sense that multiple languages are available to the student, even though the assessment may be conducted in multiple but independent languages. The second is multilingual in the sense that student's performance reveals certain practices and capabilities that characterize the use of multiple languages by multilingual users as they negotiate, mediate, or facilitate communication. Although emanating from different contexts of language education and incorporating different terms, it is possible to draw some parallels between the more recent understandings of the assessment of intercultural practices and capabilities and the notion of multilingual functioning. Studies in assessment have been undertaken in relation to the first, but, although research on actual practices of multilingual speakers has been conducted, it has not been specifically in the context of assessment. Though not explicitly foregrounded, culture/s as well as language/s is at play.

Considering the first sense of "multilingual assessment," in an 8-year system-wide study in the multilingual context of Ethiopia, Heugh et al. (2012) demonstrated the value of learning and assessment in the student's mother tongue in bi-/trilingual teaching programs. Heugh et al. (2016) draw attention to bilingual and multilingual design of large-scale, system-wide assessments of student knowledge in two or three languages, as well as the unanticipated use, on the part of students, of their bilingual or multilingual repertoires in high-stake examinations. In the research reported by Shohamy (2011), immigrant students from the former USSR and Ethiopia, when assessed in Hebrew as the language of instruction in Israeli Jewish schools,

performed less successfully than the local, native students. Such students bring prior academic and cultural knowledge to the assessment situation, but this knowledge is not captured when the assessment is conducted in a language and culture that is different from that of their primary socialization. Furthermore, as Shohamy explained, these students naturally continue to use the linguistic and cultural resources developed prior to immigration, but their capacity to use this knowledge is not assessed. In these circumstances, the picture of their multilingual and multicultural achievements is distorted.

Cenoz and Gorter (2011) also highlighted approaches that draw on the whole linguistic repertoire of multilingual speakers. They reported on an exploratory study of students' trilingual written production in Basque, Spanish, and English in schools in the Basque Country. They focused specifically on the interaction among the three languages. The study showed that consideration of writing performance across three languages revealed similar patterns in writing skills in the three languages. They also illustrated that students use multilingual practices in creative ways and that achievement is improved when practices such as codeswitching and translanguaging are employed. These practices are linguistic and also cultural.

Work in Progress

At the present stage of development, work in progress tends to be in individual, small-scale studies rather than part of large-scale programs of research and development. Conceptual work on modeling intercultural (or more precisely interlinguistic and intercultural) practices and capabilities will continue, as will consideration about the assessment of multiple languages and cultures and their relationship. Equally, discussion will continue about the assessment of capabilities beyond the linguistic (such as the capability to decenter or the capability to analyze critically or self-awareness about one's own linguistic and cultural profile). The Council of Europe's continuing work on the Common European Framework of Reference will seek to include indicators of intercultural competence because of the current desire to develop scaled, quantified levels of competence in all aspects of education. The current general education project of the Council of Europe, entitled "Competences for Democratic Culture and Intercultural Dialogue" (https://www.coe.int/t/dg4/education/descriptors_en.asp), may contribute to this line of development. Such quantification, however, runs counter to the qualitative, descriptive orientation that capturing these practices and capabilities entails.

An increasing range of research is being undertaken with a focus on multilingual functioning, especially processes such as translanguaging (Li Wei 2014; García and Li 2014). An explicit focus on the cultural and intercultural along with the linguistic and interlinguistic may add value to these research endeavors.

Some small-scale studies provide examples of work in progress. In a longitudinal study entitled "Developing English language and intercultural learning capabilities,"

Heugh (personal communication, October 2015) is incorporating translanguaging practices in the teaching, learning, and assessment of the English language of international students. The study involves practices in which students are invited to use their knowledge and expertise in their primary language in the process of developing high-level proficiency in English. Diagnostic assessment of students' written texts in Cantonese, Putonghua, and English allows for a more nuanced understanding of students' holistic capabilities in both their primary language and English (see Heugh et al. (2016)). This work is very much in line with Shohamy's (2011) desire that assessment recognizes the legitimate use and mixing of multiple languages, for it permits multilingual students to use their full linguistic, cultural, semiotic, and knowledge repertoires to interpret and create meaning. Heugh's work is demonstrating that these Chinese-speaking students also experience enhanced metalinguistic awareness of their own linguistic, cultural, and knowledge repertoire.

At the School of Oriental and African Studies, Pizziconi and Iwasaki (personal communication, October 2015) are researching the assessment of intercultural capabilities in the teaching and learning of Japanese. This work is being undertaken in the context of the AILA Research Network on Intercultural Mediation in Language and Culture Teaching and Learning. The project follows the development of linguistic and intercultural mediation capabilities in 14 learners of Japanese language before, during, and after a year of study in Japan. Through a variety of instruments, they are examining how students interpret, respond to, and negotiate identities, stereotypes, intercultural similarities and differences, the tensions arising from novel contact situations, the nature of the connections established, and how this is reflected in their language use. In short, they are investigating whether and how this long-term experience of "otherness" affects both performance and awareness.

Within the same network Angela Scarino, Anthony Liddicoat, and Michelle Kohler are developing specifications for the assessment of intercultural capabilities in languages learning in the K–12 setting in Australia. These will be used with teachers working in a range of languages to develop assessment procedures, implement them, and analyze samples of students' works for evidence of intercultural capabilities.

The new national curriculum for language learning in Australia has proposed an intercultural orientation to language teaching, learning, and assessment. Several studies related to the implementation of this curriculum, and related assessment practices, are currently being undertaken at the Research Centre for Languages and Cultures at the University of South Australia, in addition to experimenting with the design of elicitation processes.

The line of research by Cenoz and Gorter (2011) on trilingual students' participation in language practices that are shaped by the social and cultural context in the Basque Country and Friesland is continuing (see Gorter 2015) as is the work of Heugh et al. (2016).

Problems and Difficulties

In expanding the construct of communicative competence toward symbolic, intercultural, and multilingual orientations (among the many new formulations that seek to represent this expansion), there is a need for explicit consideration of peoples' situatedness in the language/s and culture/s of their primary and ongoing socialization in the distinctive contexts of linguistic and cultural diversity. This attention is central to an understanding both of culture in language assessment and the role of culture in the assessment of students' learning outside the languages of their primary socialization, in multilingual and multicultural contexts. Difficulties remain at the level of conceptualization, elicitation, and judging.

Conceptualizing Culture and Language Assessment

Further work is needed in conceptualizing the assessment of culture and the role of culture, particularly in multilingual and multicultural assessments. This may include, but is not limited to, the use of multiple languages in the assessment of content knowledge, the use of multiple languages and cultures in contemporary communication on the part of multilingual users, and a focus on interlinguistic and intercultural practices and capabilities in the assessment of additional languages. Both the conceptual work and its translation into assessment practice remain challenging because of the monolingual bias of both traditional SLA (May 2014; Leung and Scarino 2016) and traditional assessment (Shohamy 2011).

As part of this conceptual work, further consideration will need to be given to the context of culture and how it is perceived by participants in communication. Questions are being raised about the feasibility of assessing dimensions that go beyond the linguistic and the cultural, whether or not assessment philosophies and approaches can encompass the elicitation and judging of such complex practices and capabilities that go well beyond the linguistic and cultural per se, and the ethics of seeking to assess the realm of personal values, dispositions, effect, and critical awareness.

Elicitation

The traditional product orientation of assessment does not capture the processual and reflective dimensions of assessing interlinguistic and intercultural and multilingual practices and capabilities. Finding productive ways of capturing cultural and intercultural interpretations will be difficult, and, in this regard, inquiry and hermeneutic approaches are likely to be of value (Moss 2008). These would permit the capturing for the purposes of assessment not only of experiences of interlinguistic and intercultural communication but also students' understandings of and reflections on the processes of meaning-making. The use of portfolios or journals, captured over time and including reflective commentaries, would seem fruitful. The complexity of

seeking to elicit the multiple facets of interlinguistic and intercultural communication (i.e., performance, analysis, and reflection) in an integrated and holistic way remains an area for experimentation. This is an important area for language educators who are concerned with developing as well as assessing such practices and capabilities. The elicitation process is necessarily framed by some understanding of the evidence that educators might expect to see in students' performances. As the kind of evidence of this kind of language-and-culture learning goes well beyond the accuracy, fluency, appropriateness, and complexity of language use, the very nature of this evidence will also require further consideration.

Judging

As indicated earlier, there is a difficulty in judging, because of the uncertainty that arises for educators about judging student subjectivies and values. In the current state of play with assessment, what is absent is a larger frame of reference that educators need to bring to the processes of making judgments. Any instance of performance needs to be referenced against a map of other possible instances, but at this time, such a map is not available. As well, working with the notion of fixed rather than emerging criteria and scales adds complexity to the process. Educators desire certainty, when in fact there will necessarily be a great deal of uncertainty. This uncertainty relates to the absence of a shared frame of reference (such as one that they might have for a skill such as writing), but there are no firm guidelines as to what constitutes evidence. Furthermore, instances of communication of meaning across languages will be highly variable contextually, and yet it is precisely this linguistic and cultural variability and the linguistic and cultural situatedness of the participants that is being assessed in culture and language assessment.

In all three areas – conceptualizing, eliciting, and judging – the resilience of traditional practices is a major difficulty. In research, it is clear that both large-scale and smaller, grounded, ethnographic studies will be needed, focused on the assessment of interlinguistic and intercultural and multilingual practices and capabilities. It will be particularly fruitful for work in progress to be shared, compared, and theorized across research groups, given the immense diversity of local contexts of language-and-culture learning and its assessment.

Having highlighted the resilience of traditional assessment practices and their monolingual and monocultural bias, teacher education becomes a complex process of unlearning and learning. Teachers' assessment practices are heavily constrained by the requirements of the education systems in which they work. These requirements tend to be designed for accountability purposes more than for educational ones; therefore the environment is often not conducive to the kind of alternative practices that the assessment of these capabilities will require (see Scarino 2013 for a detailed discussion).

Finally, it must be recognized that this kind work in assessment, both in terms of practices and research, will be resource-intensive and raise issues of practicability. However, what is at stake in considering culture and assessment is the very nature of

language learning and its assessment and doing justice to capturing and giving value to the learning and achievements of students who are developing their multi-/interlinguistic and multi-/intercultural capabilities.

Future Directions

What is needed is a program of research, undertaken in diverse contexts, that considers the meaning-making processes of students in their multi-/interlinguistic work and multi-/intercultural work. These are likely to include processes such as decentring and translanguaging, mediating understanding across multiple languages, and paying greater attention to the positioning of students. Evidence might include analyses of moment-to-moment actions/interactions/reactions, conversations, or introspective processes that probe students' meanings; surveys, interviews, and self-reports; and reflective summaries and commentaries on actions, and reactions. Also needed is a focus on identifying and naming or describing the distinctive capabilities that can be characterized as multi-/interlinguistic and multi-/intercultural. These are the unique capabilities that bi-/multilingual students display as they move across diverse linguistic and cultural worlds. They are likely to include not only knowledge and skill but also embodied experience and their consideration of language/s and culture/s within that experience. Here it would become necessary to understand not only students' ideas but also their life worlds, their linguistic and cultural situatedness, and their histories and values; to understand the way these form the interpretive resources that they bring to the reciprocal interpretation and creation of meaning; and to understand both themselves (intraculturally) and themselves in relation to others (interculturally).

Cross-References

- ▶ Assessing English as a Lingua Franca
- ▶ Assessing Second/Additional Language of Diverse Populations
- ▶ Using Portfolios for Assessment/Alternative Assessment
- ▶ The Common European Framework of Reference (CEFR)

Related Articles in the Encyclopedia of Language and Education

Anne-Brit Fenner: Cultural Awareness in the Foreign Language Classroom. In Volume: Language Awareness and Multilingualism

Francesca Helm: Critical Approaches to Online Intercultural Language Education. In Volume: Language, Education and Technology

Yiqi (April) Liu: Popular Culture and Teaching English to Speakers of Other Languages (TESOL). In Volume: Language, Education and Technology

Brenday O'Connor, Norma González: Language Education and Culture. In Volume: Language Policy and Political Issues in Education

Robert O'Dowd: Online Intercultural Pedagogy and Language Education. In Volume: Language, Education and Technology

References

Atkinson, D. (1999). TESOL and culture. *TESOL Quarterly, 33*, 625–654. doi:10.2307/3587880.

Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox et al. (Eds.), *Language testing reconsidered* (pp. 41–71). Ottawa: Ottawa University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: 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: Oxford University Press.

Bennett, M. (1986). Towards ethnorelativism: A developmental model of intercultural sensitivity. In M. Paige (Ed.), *Cross-cultural orientation: New conceptualisation and applications* (pp. 27–70). New York: University Press of America.

Block, D. (2003). *The social turn in second language acquisition*. Washington, DC: Georgetown University Press.

Block, D. (2014). Moving beyond 'lingualism': Multilingual embodiment and multimodality in SLA. In S. May (Ed.), *The multilingual turn: Implications for SLA, TESOL and bilingual education* (pp. 54–77). New York/London: Routledge.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Clevedon: Multilingual Matters.

Byram, M. (2014). Twenty-five years on – From cultural studies to intercultural citizenship. *Language, Culture and Curriculum, 27*, 209–225. doi:10.1080/07908318.2014.974329.

Byram, M., & Zarate, G. (1994). *Définitions, objectifs et évaluation de la compétence socio-culturelle*. Strasbourg: Report for the Council of Europe.

Byrnes, H. (2010). Revisiting the role of culture in the foreign language curriculum. Perspectives column. *Modern Language Journal, 94*, 315–336. doi:10.1111/j.1540-4781.2010.01023.x.

Canale, M., & Swain, M. (1980). A domain description for core FSL: Communication skills. In Ontario Ministry of Education (Ed.), *The Ontario assessment instrument pool: French as a second language, junior and intermediate divisions* (pp. 27–39). Toronto: Ontario Ministry of Education.

Canale, M., & Swain, M. (1981). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1–47.

Cenoz, J., & Gorter, D. (2011). Focus on multilingualism: A study of trilingual writing. *Modern Language Journal, 95*, 356–369. doi:10.1111/j.1540-4781.2011.01206.x.

Chalhoub–Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing, 20*, 369–383. doi:10.1191/0265532203lt264oa.

Deardorff, D. (2009). Implementing intercultural competence assessment. In D. Deardorff (Ed.), *The Sage handbook of intercultural competence* (pp. 477–491). Los Angeles: Sage.

Delandshere, G. (2002). Assessment as inquiry. *Teachers College Record, 104*, 1461–1484.

García, O., & Li, W. (2014). *Translanguaging: Language, bilingualism and education*. Basingstoke: Palgrave Macmillan.

Geertz, C. (2000). *Available light: Anthropological reflections on philosophical topics*. Princeton: University Press.

Gorter, D. (2015). Multilingual interaction and minority languages: Proficiency and language practices in education and society. *Language Teaching, 48*, 83–98. doi:10.1017/S0261444812000481.

Gumperz, J. J., & Levinson, S. C. (Eds.). (1996). *Rethinking linguistic relativity.* Cambridge: Cambridge University Press.

Gutierrez, K., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher, 32*(5), 19–25. doi:10.3102/0013189X032005019.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education, 4*, 93–116.

Halliday, M. A. K. (1999). The notion of 'context' in language education. In M. Ghadessy (Ed.), *Text and context in functional linguistics* (pp. 1–24). Philadelphia: Benjamins.

Heugh, K., Benson, C., Yohannes, M. A. G., & Bogale, B. (2012). Implications for multilingual education: Student achievement in different models of education in Ethiopia. In T. Skutnabb-Kangas & K. Heugh (Eds.), *Multilingual education and sustainable diversity work: From periphery to centre* (pp. 239–262). New York/London: Routledge.

Heugh, K., Prinsloo, C., Makgamantha, M. M., Diedericks, G., & Winnaar, L. (2016). Multilingualism(s) in system-wide assessment: A politics of innovation and resistance. Special issue on 'Multilingual Assessment'. *Language and Education.*

Heugh, K., Li, X., & Song, Y. (2016). Multilingualism and translanguaging in the teaching of and through English: Rethinking linguistic boundaries in an Australian university. In I. Walkinshaw, B. Fenton-Smith & P. Humphries (Eds.), *English medium instruction in higher education in the Asia-Pacific. From policy to pedagogy.* Dordrecht: Springer.

Houghton, S. A. (2013). Making intercultural communicative competence and identity-development visible for assessment purposes in foreign language education. *Language Learning Journal, 41*, 311–325. doi:10.1080/09571736.2013.836348.

Kramsch, C. (1986). From language proficiency to interactional competence. *Modern Language Journal, 70*, 366–372.

Kramsch, C. (1993). *Context and culture in language teaching*. Oxford: Oxford University Press.

Kramsch, C. (2004). Language, thought and culture. In A. Davies & C. Elder (Eds.), *Handbook of applied linguistics* (pp. 235–261). Malden: Blackwell Publishing.

Kramsch, C. (2006). From communicative competence to symbolic competence. *Modern Language Journal, 90*, 249–252. doi:10.1111/j.1540-4781.2006.00395_3.x.

Kramsch, C. (2009). Discourse, the symbolic dimension of Intercultural Competence. In A. Hu & M. Byram (Eds.), *Interkulturelle Kompetenz und Fremdsprachliches Lernen. Modelle, empirie, evaluation* [Intercultural competence and foreign language learning: Models, empiricism, assessment] (pp. 107–122). Tűbingen: Gunter Narra Verlag.

Kramsch, C. (Ed.) (2014). Teaching foreign languages in the era of globalization. *The Modern Language Journal*, *98*(Special issue 1), 296.

Lee, C. D. (2008). The centrality of culture to the scientific study of learning and development: How an ecological framework in education research facilitates civic responsibility. *Educational Researcher, 37*, 267–279. doi:10.3102/0013189X08322683.

Leung, C., & Scarino, A. (2016). Reconceptualising the nature and outcomes in language/s education. *The Modern Language Journal, 100*(Suppl 2016). doi:10.111/modl.123000026-7902/16/81-95.

Liddicoat, A. J., & Scarino, A. (2013). *Intercultural language teaching and learning*. Malden: Wiley-Blackwell.

Lussier, D., Ivanus, D., Chavdarova-Kostova, S., Golubina, K., Skopinskaja, L., Wiesinger, S., & de la Maya Retamar, G. (2007). Guidelines for the assessment of intercultural communicative competence. In I. Lazar, M. Huber-Kriegler, D. Lussier, G. S. Matei, & C. Peck (Eds.), *Developing and assessing intercultural communicative competence: A guide for language teachers and teacher educators* (pp. 23–39). Strasbourg/Graz: European Centre for Modern Languages and Council of Europe Publishing. http://archive.ecml.at/mtp2/publications/B1_ICCinTE_E_internet.pdf. Accessed 15 Nov 2015.

May, S. (Ed.). (2014). *The multilingual turn: Implications for SLA, TESOL and bilingual education*. New York/London: Routledge.

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden: Blackwell Publishing.

Moss, P. A. (2008). Sociocultural implications for assessment: Classroom assessment. In P. A. Moss, D. C. Pullin, P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity and opportunity to learn* (pp. 222–258). Cambridge: Cambridge University Press.

Risager, K. (2007). *Language and culture pedagogy: From a national to transnational paradigm*. Clevedon: Multilingual Matters.

Sapir, E. (1962). In D. Mandelbaum (Ed.), *Culture, language and personality: Selected essays*. Berkeley: University of California Press.

Scarino, A. (2010). Assessing intercultural capability in learning languages: A renewed understanding of language, culture, learning and the nature of assessment. *Modern Language Journal, 94*(2), 324–329. doi:10.1111/j.1540-4781.2010.01026.x.

Scarino, A. (2013). Language assessment literacy as self-awareness. Understanding the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(Special issue 3), 309–327. doi:10.1177/0265532213480128.

Sercu, L. (2004). Assessing intercultural competence: A framework for systematic test development in foreign language education and beyond. *Intercultural Education, 15*, 73–89. doi:10.1080/1467598042000190004.

Shohamy, E. (2011). Assessing multilingual competences: Adopting construct valid assessment policies. *Modern Language Journal*, *95*, 418–429. doi:10.1111/j.1540-4781.2011.01210.x.0026-7902/11/418–429.

Steffensen, S. V., Michiko, U., & Kramsch, C. (2014). The ecology of intercultural interaction: Timescales, temporal ranges and identity dynamics. *Language Sciences, 41A*, 41–59. doi:10.1016/j.langsci.2013.08.006.

Wei, L. (2014). Who's teaching whom? Co-learning in multilingual classrooms. In S. May (Ed.), *The multilingual turn: Implications for SLA, TESOL and bilingual education* (pp. 167–190). New York/London: Routledge.

Whorf, B. L. (1940/1956). In J. B. Carroll (Ed.), *Language, thought and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.