

Educational Linguistics

Volume 26

Series Editor

Francis M. Hult, Lund University, Sweden

Editorial Board

Marilda C. Cavalcanti, Universidade Estadual de Campinas, Brazil

Jasone Cenoz, University of the Basque Country, Spain

Angela Creese, University of Birmingham, United Kingdom

Ingrid Gogolin, Universität Hamburg, Germany

Christine Hélot, Université de Strasbourg, France

Hilary Janks, University of Witwatersrand, South Africa

Claire Kramsch, University of California, Berkeley, U.S.A

Constant Leung, King's College London, United Kingdom

Angel Lin, University of Hong Kong, Hong Kong

Alastair Pennycook, University of Technology, Sydney, Australia

Educational Linguistics is dedicated to innovative studies of language use and language learning. The series is based on the idea that there is a need for studies that break barriers. Accordingly, it provides a space for research that crosses traditional disciplinary, theoretical, and/or methodological boundaries in ways that advance knowledge about language (in) education. The series focuses on critical and contextualized work that offers alternatives to current approaches as well as practical, substantive ways forward. Contributions explore the dynamic and multi-layered nature of theory-practice relationships, creative applications of linguistic and symbolic resources, individual and societal considerations, and diverse social spaces related to language learning.

The series publishes in-depth studies of educational innovation in contexts throughout the world: issues of linguistic equity and diversity; educational language policy; revalorization of indigenous languages; socially responsible (additional) language teaching; language assessment; first- and additional language literacy; language teacher education; language development and socialization in nontraditional settings; the integration of language across academic subjects; language and technology; and other relevant topics.

The *Educational Linguistics* series invites authors to contact the general editor with suggestions and/or proposals for new monographs or edited volumes. For more information, please contact the publishing editor: Jolanda Voogd, Associate Publishing Editor, Springer, Van Godewijkstraat 30, 3300 AA Dordrecht, The Netherlands.

More information about this series at <http://www.springer.com/series/5894>

Martin East

Assessing Foreign Language Students' Spoken Proficiency

Stakeholder Perspectives
on Assessment Innovation

 Springer

Martin East
Faculty of Education and Social Work
The University of Auckland
Auckland, New Zealand

ISSN 1572-0292

ISSN 2215-1656 (electronic)

Educational Linguistics

ISBN 978-981-10-0301-1

ISBN 978-981-10-0303-5 (eBook)

DOI 10.1007/978-981-10-0303-5

Library of Congress Control Number: 2015960962

Springer Singapore Heidelberg New York Dordrecht London

© Springer Science+Business Media Singapore 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer Science+Business Media Singapore Pte Ltd. is part of Springer Science+Business Media
(www.springer.com)

Foreword

The ability to use a language for spoken communication is one of the main reasons many people study a foreign language, and learners often evaluate their success in language learning, as well as the effectiveness of their language course, on the basis of how well they feel they have improved in their spoken-language proficiency. When a foreign language is an assessed school subject, the procedures used to arrive at a valid account of learners' ability to speak is of crucial importance to schools, teachers and learners. However, the assessment of spoken-language proficiency, or 'speaking skills', has been somewhat problematic in the history of language teaching. On the one hand, the construct of spoken-language proficiency itself has sometimes been inadequately theorised. Whereas 'speaking skills' covers a wide range of different modes of discourse, including small talk, casual conversations, telephone conversations, transactions, discussions, interviews, meetings, presentations and debates, conventional approaches to oral proficiency assessment have often assumed that performance on an oral interview can be assumed to represent general spoken-language ability rather than simply ability related to one mode of discourse. In reality, each genre has distinct features and characteristics and each poses quite different issues for teaching, learning and assessment.

In addition the practical logistics of conducting oral assessments make assessment of authentic language use difficult. Unlike the assessment of other language skills, speaking ability can only be observed in face-to-face interaction. However, the kind of interaction that results may be influenced by many factors, such as those related to the context, the interlocutors, the setting and the task. A proficiency test that is part of a high-stakes school-based assessment is very different from language used outside of the classroom in a naturalist context for communication. It is challenging for proficiency tests not only to replicate but also to capture authentic language use.

A further issue results from the fact that changes in assessment practices – such as in the New Zealand example articulated in this book, that is, a move from assessment *of* learning towards assessment *for* learning – may not align with learners'

expectation about the nature and function of assessment. If teachers move away from summative tests and adopt continuous assessment based on an assessment *for* learning approach, there may be resistance from learners, since what counts for them may be their end of course grades rather than the teacher's well-intentioned philosophy of assessment.

Against this background, teachers, curriculum developers and language teaching professionals will find the current book a unique and invaluable account of how issues such as those mentioned above were addressed in the context of a large-scale curriculum innovation in language teaching in New Zealand. The context is the teaching of foreign languages in New Zealand secondary schools and the introduction of a new curriculum that seeks to achieve a learner-based rather than teacher-dominated approach to learning. In applying this philosophy to foreign language teaching, an attempt was made to develop a new approach to oral proficiency assessment. A one-off summative assessment based on students' performances on an end-of-year interview has been replaced with a series of peer-to peer interactions that seek to provide learners with opportunities to show what they can do with a language (rather than what they know about it) and how effectively they can use it for authentic communication. This book results from a 2-year study of how the new assessment approach worked in practice, as experienced by both teachers and students.

In this book Martin East provides a fascinating account of how the new assessment approach was introduced; how it differs from previous approaches to assessment; the impact of the approach on teachers, teaching and learners; and the issues it created for teachers, learners and schools. The importance of this book derives from a number of features:

- It is a model case study of curriculum innovation in foreign language education.
- It provides an account of an approach to validity that contrasts a standard psychometric approach derived from performance scores with one that reflects stakeholders' views of the innovation.
- It reflects an approach in which assessment is designed to be an integral part of the teaching and learning programme and that makes use of activities that are typical in foreign language classrooms.
- Rather than employing a one-off interview, it makes use of a series of tasks to capture the dynamic and interactive nature of spoken interaction.
- It makes use of qualitative methods to capture the subjective experiences of teachers and students.
- It provides a detailed account of both the perceived benefits and the perceived disadvantages of the innovation.
- It raises issues that are of much wider significance than the specific context (New Zealand students learning foreign languages) in which they were studied.

Assessing Foreign Language Students' Spoken Proficiency makes an outstanding and original contribution to the field of second and foreign language teaching, providing a theory and research-based account of the development of a learner-centred approach to oral proficiency assessment. It is an important resource for teachers and teacher educators as well as assessment and curriculum specialists worldwide. It deserves to be widely read.

January 2016

Jack C. Richards

Acknowledgments

This book represents the culmination of several years of research work which I would not have been able to achieve without a range of help and support. First and foremost, I would like to acknowledge the funding I received from the University of Auckland in the form of a Research Excellence Award (ECREA 3701329). This funding provided the opportunity both to administer the national teacher survey and to travel across the country to carry out interviews with teachers. The award also funded transcription and data analysis costs, as well as opportunities to present emerging findings from the research in a range of international fora.

Additionally, I was able to engage the services of two summer scholars who worked with me at different stages of the research process. I acknowledge Joseph Poole who undertook some of the transcribing and data entry, alongside initial coding of the teacher and student surveys, and Abby Metson, who undertook initial coding of themes emerging from the interviews.

I thank my wife, Julie, whose initial conversations with me helped to shape the scoping of the project and whose statistical expertise contributed to the quantitative analyses I report.

I am grateful to Professor Jack C Richards, an internationally recognised authority in second and foreign language teaching and learning, for his interest in my work and his willingness to write the foreword.

My thanks to Nick Melchior, Senior Editor (Education) for Australia/New Zealand at Springer, for his enthusiasm to see this research published as part of Springer's *Educational Linguistics* series.

Bearing in mind the duration of this project, it is inevitable that emerging findings have been published in other fora. Two articles have published aspects of the data I report in Chaps. 5 and 6:

East, M. (2014). Working for positive outcomes? The standards-curriculum alignment for *Learning Languages* and its reception by teachers. *Assessment Matters*, 6, 65–85.

East, M. (2015). Coming to terms with innovative high-stakes assessment practice: Teachers' viewpoints on assessment reform. *Language Testing*, 32(1), 101–120. <http://dx.doi.org/10.1177/0265532214544393>

Additionally, two articles that informed the background material I report in Chap. 3 are:

East, M., & Scott, A. (2011a). Assessing the foreign language proficiency of high school students in New Zealand: From the traditional to the innovative. *Language Assessment Quarterly*, 8(2), 179–189. <http://dx.doi.org/10.1080/15434303.2010.538779>

East, M., & Scott, A. (2011b). Working for positive washback: The standards-curriculum alignment project for *Learning Languages*. *Assessment Matters*, 3, 93–115.

I would like to thank Adèle Scott for her support in verifying the accuracy of the historical account of events I present in Chap. 3.

Last, and certainly by no means least, I thank sincerely all the participants in this project who completed a survey, or who asked their students to complete a survey, or who participated in an interview with me. Without the help and support of those who are willing to give their time when invited to do so, research projects such as the one reported here cannot come to fruition. The willingness of participants enables their voices to be represented.

Contents

1 Mediating Assessment Innovation: Why Stakeholder Perspectives Matter	1
1.1 Introduction.....	1
1.2 Background: The Importance of Interaction in Foreign Languages Pedagogy	3
1.2.1 Communicative Language Teaching	3
1.2.2 Communicative Language Testing	5
1.3 Curriculum and Assessment Reforms in New Zealand.....	6
1.3.1 Overview.....	6
1.3.2 Implementing Assessment Reform: A Risky Business.....	7
1.4 Assessment Validation.....	8
1.4.1 Fundamental Considerations	8
1.4.2 The Contribution of Assessment Score Evidence to a Validity Argument.....	10
1.4.3 The Limitations of Assessment Score Evidence to a Validity Argument	12
1.4.4 Towards a Broader Understanding of Assessment Validation	13
1.4.5 A Qualitative Perspective on Assessment Validation	16
1.5 The Structure of This Book.....	18
1.6 Conclusion.....	19
References.....	21
2 Assessing Spoken Proficiency: What Are the Issues?	25
2.1 Introduction.....	25
2.2 What Does It Mean to Communicate Proficiently?	26
2.2.1 Communicative Competence as the Underlying Theoretical Framework.....	26
2.2.2 Developing the Framework of Communicative Competence	27

2.3	Static or Dynamic.....	31
2.3.1	The Static Assessment Paradigm.....	32
2.3.2	The Dynamic Assessment Paradigm.....	33
2.3.3	Static or Dynamic – A Complex Relationship.....	35
2.4	Task-Based or Construct Based.....	37
2.4.1	The Centrality of the Task.....	37
2.4.2	The Importance of the Construct.....	39
2.5	Single or Paired Performances.....	40
2.5.1	Single Performance Assessments.....	41
2.5.2	Paired/Group Performance Assessments.....	42
2.6	Conclusion.....	45
	References.....	46
3	Introducing a New Assessment of Spoken Proficiency: <i>Interact</i>	51
3.1	Introduction.....	51
3.2	The New Zealand Landscape for Assessment – A Shifting Environment.....	52
3.2.1	The 1990s: A Mismatch Between Curricular Aims and High-Stakes Assessment.....	53
3.2.2	The NCEA System: The Beginnings of Reform.....	56
3.2.3	The Impact of Assessment Mismatch on FL Programmes.....	57
3.2.4	The NCEA for Languages – 2002–2010.....	59
3.3	Towards a Learner-Centred Model for High-Stakes Assessment.....	61
3.3.1	2007: The Advent of a New Curriculum.....	61
3.3.2	NCEA Mark II.....	63
3.4	Revising the Assessments for Languages.....	64
3.4.1	2008: The First SCALES Meeting.....	64
3.4.2	2009: The Second SCALES Meeting.....	66
3.4.3	2010: A Further Opportunity to Confirm the New Assessments.....	67
3.4.4	2011 Onwards: Support for the Implementation of <i>Interact</i>	69
3.5	Conclusion.....	72
	References.....	73
4	Investigating Stakeholder Perspectives on <i>Interact</i>	77
4.1	Introduction.....	77
4.2	Bachman and Palmer’s Test Usefulness Framework.....	78
4.2.1	Construct Validity and Reliability.....	79
4.2.2	Interactiveness, Impact, Practicality and Authenticity.....	81
4.3	2011 Onwards: <i>Interact</i> in Practice.....	84
4.4	The Theoretical Usefulness of <i>Interact</i>	87
4.5	A Study into Teachers’ and Students’ Views.....	88
4.6	Study Stage I.....	90
4.6.1	Nationwide Teacher Survey.....	90
4.6.2	Piloting the Teacher Survey.....	91

4.6.3	Administering the Main Survey	92
4.6.4	Teacher Interviews.....	94
4.7	Stage II.....	95
4.7.1	Teacher Interviews.....	95
4.7.2	Student Surveys	95
4.8	Conclusion.....	97
	References.....	98
5	The Advantages of <i>Interact</i>	101
5.1	Introduction.....	101
5.2	The Nationwide Teacher Survey – Section I.....	102
5.2.1	Overview.....	102
5.2.2	Perceived Relative Usefulness of <i>Converse</i> and <i>Interact</i>	103
5.2.3	Variations in Teacher Responses	106
5.2.4	Differences in Perception According to Principal Language Taught.....	108
5.2.5	Differences in Perception According to Whether or Not Using <i>Interact</i>	110
5.3	Advantages of <i>Interact</i> – Survey Data	111
5.3.1	Authenticity and Interactiveness.....	112
5.3.2	Positive Impact	114
5.3.3	Validity, Reliability and Potential for Washback	116
5.4	Advantages of <i>Interact</i> – Interviews	117
5.4.1	Authenticity and Interactiveness.....	118
5.4.2	Positive Impact	120
5.4.3	Validity, Reliability and Potential for Washback	122
5.5	Conclusion.....	123
	References.....	124
6	The Disadvantages of <i>Interact</i> and Suggested Improvements	125
6.1	Introduction	125
6.2	Disadvantages of <i>Interact</i> – Survey Data.....	126
6.2.1	Impracticality.....	126
6.2.2	Negative Impact – Unrealistic Expectations.....	129
6.2.3	Negative Impact – Interlocutor Variables	130
6.3	Suggestions for Improvement – Survey Data.....	130
6.3.1	Reduce the Number of Interactions Required	131
6.3.2	Allow Provision for Scaffolding/Rehearsal.....	132
6.3.3	Provide More Examples and More Flexible Options	134
6.4	Disadvantages of <i>Interact</i> – Interviews.....	135
6.4.1	Impracticality.....	135
6.4.2	Negative Impact – Too Much Work for What It Is Worth	137
6.4.3	Negative Impact – Interlocutor Variables	138
6.4.4	The Challenges of ‘Spontaneous and Unrehearsed’	139

6.5	Suggestions for Improvement – Interviews.....	140
6.5.1	Clarifying ‘Spontaneous and Unrehearsed’.....	141
6.5.2	The Task is Everything.....	143
6.6	Conclusion.....	144
	References.....	145
7	Interact and Higher Proficiency Students:	
	Addressing the Challenges	147
7.1	Introduction.....	147
7.2	Examples of Task Types.....	149
7.2.1	Talking About the Environment.....	150
7.2.2	Mariage Pour Tous.....	153
7.2.3	Cat Café.....	153
7.2.4	Getting Students to Take the Lead.....	155
7.3	Problems Emerging.....	156
7.3.1	Spontaneous and Unrehearsed.....	156
7.3.2	Moving Away from Grammar.....	159
7.4	Back to the Task.....	162
7.5	Conclusion.....	165
	References.....	166
8	Interact and Higher Proficiency Students:	
	Concluding Perspectives	167
8.1	Introduction.....	167
8.2	Working for Washback.....	168
8.3	The Student Surveys.....	171
8.3.1	Section I.....	172
8.3.2	Taking a Closer Look at the Numbers.....	173
8.4	Student Survey Responses – <i>Converse</i>	175
8.5	Student Survey Responses – <i>Interact</i>	177
8.5.1	Spontaneity Versus Grammar.....	178
8.5.2	Types of Task.....	182
8.5.3	Peer-to-Peer Interactions.....	183
8.5.4	Working for Washback.....	184
8.6	Conclusion.....	185
	References.....	187
9	Coming to Terms with Assessment Innovation: Conclusions	
	and Recommendations	189
9.1	Introduction.....	189
9.2	Theoretical Underpinnings of <i>Interact</i>	190
9.3	Summary of Findings.....	192
9.3.1	Overview.....	192
9.3.2	Positive Dimensions of Assessments Such as <i>Interact</i>	193
9.3.3	Negative Dimensions of Assessments Such as <i>Interact</i>	195

- 9.4 Static or Dynamic: A Fundamental Problem 196
 - 9.4.1 Is *Interact* a Test?..... 196
 - 9.4.2 What Do We Want to Measure?..... 199
- 9.5 Where to from Here?..... 201
 - 9.5.1 Scenario 1 202
 - 9.5.2 Scenario 2 202
- 9.6 Recommendations 204
- 9.7 Limitations and Conclusion 206
- References 209

- Bibliography** 213

- Index**..... 225

List of Figures

Fig. 3.1	The original NCEA assessment matrix.....	60
Fig. 3.2	The revised NCEA assessment matrix.....	70
Fig. 3.3	Key changes between <i>converse</i> and <i>interact</i>	71
Fig. 4.1	Outcome requirements of interactions	85
Fig. 4.2	Procedure for eliciting strength of perception	90
Fig. 5.1	Numbers of survey respondents (<i>left</i>) compared to numbers of NCEA (senior secondary) students (2012) (<i>right</i>)	102
Fig. 5.2	Numbers of survey respondents using/not using <i>interact</i>	103
Fig. 5.3	Percentage histogram of difference scores (<i>converse</i> – <i>interact</i>) by measure	107
Fig. 5.4	Difference scores averaged across constructs	108
Fig. 5.5	Sub-construct differences in mean (<i>converse</i> v. <i>interact</i>) by language taught	109
Fig. 5.6	Sub-construct differences in mean (<i>converse</i> v. <i>interact</i>) by whether or not using <i>interact</i>	110
Fig. 8.1	Student survey mean responses by measure (<i>converse</i> v. <i>interact</i>).....	173
Fig. 8.2	<i>Converse</i> – range of responses by measure.....	174
Fig. 8.3	<i>Interact</i> – range of responses by measure.....	174

List of Tables

Table 3.1	Grades and percentage equivalents (School C and Bursary)	55
Table 4.1	Stages of the study	89
Table 4.2	Taxonomy of emerging themes from the survey, Section II	93
Table 5.1	Overall means and differences in means (teachers): <i>converse</i> and <i>interact</i>	104
Table 5.2	Differences in standardised means between <i>converse</i> and <i>interact</i>	105
Table 5.3	Analyses of variance of difference scores for each sub-construct by use of <i>interact</i>	111
Table 5.4	Frequencies of mentioning advantages of <i>interact</i>	112
Table 5.5	Interview participants (Stage I).....	117
Table 6.1	Frequencies of mentioning disadvantages of <i>interact</i>	126
Table 6.2	Frequencies of mentioning improvements to <i>interact</i>	131
Table 7.1	Interview participants (Stage II).....	149
Table 8.1	Overall means and differences in means (students): <i>converse</i> and <i>interact</i>	172
Table 8.2	Student survey participants (Stage II)	178