Testing the Spoken English of Young Norwegians

A study of test validity and the role of ‘smallwords’ in contributing to pupils’ fluency
Testing the Spoken English of Young Norwegians

A study of test validity and the role of ‘smallwords’ in contributing to pupils’ fluency

Angela Hasselgreen
I Oral language assessment and conversation analysis

Also in this series:

An investigation into the comparability of two tests of English as a Foreign Language: The Cambridge-TOEFL comparability study
Lyle F. Bachman, F. Davidson, K. Ryan, I.-C. Choi

Test taker characteristics and performance: A structural modeling approach
Antony John Kunnan

Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem
Michael Milanovic, Nick Saville

The development of IELTS: A study of the effect of background knowledge on reading comprehension
Caroline Margaret Clapham

Verbal protocol analysis in language testing research: A handbook
Alison Green

A multilingual glossary of language testing terms
Prepared by ALTE members

Dictionary of language testing
Alan Davies, Annie Brown, Cathie Elder, Kathryn Hill, Tom Lumley, Tim McNamara

Learner strategy use and performance on language tests: A structural equation modelling approach
James Enos Purpura

Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida
Antony John Kunnan

Issues in computer-adaptive testing of reading proficiency
Micheline Chalhoub-Deville

Experimenting with uncertainty: Essays in honour of Alan Davies
A. Brown, C. Elder, N. Iwashita, E. Grove, K. Hill, T. Lumley, K. O’Loughlin, T. McNamara

An empirical investigation of the componentiality of L2 reading in English for academic purposes
Cyril Weir

The equivalence of direct and semi-direct speaking tests
Kieran O’Loughlin

A qualitative approach to the validation of oral language tests
Anne Lazaraton
1 Oral language assessment and conversation analysis

Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002
Edited by Cyril Weir and Michael Milanovic

European language testing in a global context
Edited by Cyril Weir and Michael Milanovic

Unpublished

A Modular Approach to Testing English Language Skills: The development of the Certificates in English Language Skills (CELS) examinations:
Roger Hawkey

Changing language teaching through language testing: A washback study
Liying Cheng

The Impact of High-Stakes Examinations on Classroom Teaching: A Case Study Using Insights from Testing and Innovation Theory
Dianne Wall
# Contents

## Acknowledgements  
[Page xii]

Series Editor’s notes [Page xiii]

## Chapter 1  
**Introduction** [Page 1]  
- Test validation [Page 1]  
- Fluency and smallword use [Page 2]  
- The test [Page 3]  
- Research questions [Page 3]  
- Data and methods [Page 4]  
- Organisation of the book [Page 5]

## PART ONE: TEST VALIDATION  

## Chapter 2  
**Test validation** [Page 9]  
- Validation – an overview [Page 10]  
- Content validation [Page 12]  
- Face validation [Page 13]  
- Response validation [Page 14]  
- Washback validation [Page 17]  
- Consequential validation [Page 18]  
- Criterion-related validation [Page 18]  
- Reliability [Page 20]  
- Test bias [Page 22]  
- Construct validation [Page 23]  
  - The narrower view of construct validity [Page 24]  
  - The broader, unifying, view of construct validity [Page 25]  
- Threats to validity summarised [Page 27]  
- A unified framework for validation [Page 28]  
- Six central aspects of validity [Page 29]  
- A validation framework [Page 30]  
- Towards the validation process [Page 32]
Contents

Chapter 3
Communicative language ability 33
  Towards a model of communicative competence 34
  Models of communicative competence reviewed 35
  A suitable model of CLA 39
  Describing the domain of CLA 43
    Speaking 43
    The situation of the testees 46
  Operationalising components of CLA 49
    Operationalising microlinguistic ability 50
    Operationalising textual ability 51
    Operationalising pragmatic ability 52
    Operationalising strategic ability 53
  Some conclusions on CLA and the significance of smallwords 54
    Summary 55

Chapter 4
Validation of the test ‘as it stands’ 58
  The aims and purposes of the EVA testing 59
  Speaking test specifications 59
    Specifications for elicitation procedures 60
    Specifications for scoring procedures 62
  The validation process 65
    The CONTENT aspect of validity 66
    The SUBSTANTIVE aspect of validity 71
    The STRUCTURAL aspect of validity 74
    The GENERALISABILITY aspect of validity 83
    The EXTERNAL aspect of validity 84
    The CONSEQUENTIAL aspect of validity 86
  Summary and conclusions 88
    Conclusion on the extent to which the model of CLA is represented in the test 88
    Conclusions on the validity of the test 90

Chapter 5
Validation based on scoring data 96
  Data and methods 96
  The a posteriori validation process 99
    The EXTERNAL aspect 100
    The CONTENT aspect: test bias with respect to gender 103
      Generalisability 104
        Inter-rater reliability 104
PART TWO: FLUENCY AND SMALLWORD USE

Chapter 6
Fluency and smallwords – making the connection

Fluency
Pinning down fluency
Identifying elements of fluency
A language of fluency?
Fluency summarised
Forging a link between smallwords and fluency
Smallwords in other people’s books
Smallwords and fluency in relevance theory terms
The essence of relevance theory
Proposing a role for smallwords in relevance theory
The work of smallwords in optimising fluency
Levett’s perspective: speech production and fluency
A framework for analysing smallword signals
Summary

Chapter 7
Smallwords and other fluency markers: quantitative analysis

The approach
The data
The smallwords
Hypotheses and research questions
Method
Findings on temporal variables
Filled pausing
Mean length of turn
Conclusions on temporal variables
Findings on smallwords
General smallword use: quantity and distribution
Range and variety in smallword use
Smallword use summed up
Smallwords and filled pausing
Summary
Chapter 8
The signalling power of smallwords 183

The approach 184
Data, hypotheses and research questions 185
Method 185
Defining and analysing evidence that smallwords are used to send signals 188
Expressing the communicative intention 188
Signalling whether the speaker intends to take, hold or yield the turn 189
Signalling an oblique response 192
Pointing to the context for interpretation 194
Signalling a break with the initial context created by the previous speaker ('mode changing') 194
Signalling a mid-utterance break with context created by the speaker’s own immediately preceding speech 196
Indicating the cognitive effect of the previous utterance 200
Signalling a cognitive change of state, resulting from the previous utterance 201
Indicating the degree of vagueness or commitment: Signalling a softening of the impact of the message, or ‘hedging’ 204
Learner-favoured hedges 208
Learner-underused hedges 209
Conclusions on hedging 213
Indicating the state of success of communication 213
Signalling the acknowledgement of smooth communication 213
Signalling an appeal to the listener to confirm or assist smooth communication 216
Summary 218

Chapter 9
The smallword user 224

Variation in smallword use 224
Gender 224
Task 226
The acquisition of smallwords 229
The implications of the findings for language education 232
Implications for assessment 233
Implications for teaching and learning 237
Summary 239
Contents

CONCLUSION 241

Chapter 10
Conclusion 243
The research questions 243
The findings 244
Theoretical findings 244
Empirical findings 248
A small word in conclusion 254

Glossary 255

References 259

Appendices 267

Index 295
Acknowledgements

This book – and the study it reports on – could never have proceeded as smoothly as it did without the support of a lot of people. First, I would like to thank my family – close and extended – and my good friends, who always supported me and never complained that I had so little time for them. Two of my sons, Nicholas and John, laboured long on transcribing students’ speech, and Nicholas put in sterling work helping assign signals to smallwords. I also owe a great debt of gratitude to Anna-Brita Stenström, first supervisor, then friend, whose astute eyes and ears were available throughout the study, and who gave so much of her own time, so generously. And I must thank Charles Alderson, whose brisk, pertinent e-mail comments opened my eyes to so much that shaped the study. I must also mention Trude Bungum, who sadly died; she made sharing an office a pleasure, and was a true and wise friend. And finally, I would like to thank Sari Luoma for letting me pick her brains during our stay in Lancaster together, in return for the loan of my bike.

Angela Hasselgreen
To improve test fairness we need an agenda for reform, which sets out clearly the basic minimum requirements for sound testing practice. Stakeholders in the testing process, in particular students and teachers, need to be able to ask the right questions of any examinations commercial or classroom based. Examination providers should be able and required to provide appropriate evidence in response to these questions.

It is now axiomatic that a test should be constructed on an explicit specification, which addresses both the cognitive and linguistic abilities involved in the language use of interest, as well as the context in which these abilities are to be performed (theory based validity and content validity). A particular administration of a test may fulfill the requirements of both these validities to a greater or lesser extent.

Next in the implementation stage when the test has been administered, we need to look at the data generated and apply statistical analyses to these to tell us the degree to which we can depend on the results (reliabilities).

Finally we can collect data on events after the test has been developed and administered (concurrent and consequential validities) to shed further light on the well foundedness of the inferences we are making about underlying abilities on the basis of test results. The focus here is on the value of the test for end users of the information provided and the extent to which such use can be justified. This takes us into the area of concurrent validity evidence where a test is measured against other external measures of the construct, and also that of consequential validity where the impact of the test on society and individuals is investigated. This consideration implies that validity does not just reside in the test itself or rather in the scores on the test but also in the inferences that are made from them.

In Chapter 2 of this volume Hasselgreen provides a clear exposition of the nature of test validation and offers a comprehensive working framework for the validation of a spoken language test. The reader is also referred to Volume 15 of this series where the operational procedures for test validation adopted by Cambridge ESOL in terms of Validity, Reliability, Impact and Practicality (VRIP) are described. It is interesting to compare the extent to which Hasselgreen’s broad conceptualisation of this area matches that of Cambridge ESOL’s operationalisation of these VRIP categories. Together they provide a solid grounding for any future work in this area.
In Chapter 3 she examines in detail how communicative language ability (CLA), a central element of a test’s theory based validity, might be operationalised in the evaluation of the Norwegian speaking test, for lower secondary school students of English (EVA). As such it represents one of the few reported attempts to operationalise Bachman’s seminal cognitive model of language ability.

In Chapters 4 and 5 she takes the broader validation framework developed in Chapter 2 and applies it to the EVA test and so provides test developers with a working example of how validation might be done in practice. She was able to evaluate all aspects of communicative competence in EVA as it had been defined in the literature to date. Published studies of this type are regretfully rare in the testing literature and Hasselgreen’s case study illuminates this vital area of our field in an accessible well written account of a validation carried out on this spoken language test in Norway.

Her validation of the existing test system throws up serious problems in the scoring instruments. In particular the band scale relating to fluency does not adequately account for the aspects of CLA measured by the test particularly as regards textual and strategic ability because it lacks explicit reference to the linguistic devices that contribute to fluency. Low inter-rater correlations on message and fluency discussed in Chapter 5 in the discussion of a posteriori validation based on test scores further points to the problem of vagueness in the existing definitions of these criteria. This provides the link to the second part of the monograph; how to establish ‘more specific, unambiguous, data-informed ways of assessing fluency’. As such it addresses the emerging consensus that rating scale development should be data driven.

In Part 2 Hasselgreen accordingly focuses on one aspect of the validation framework that frequently generates much discussion in testing circles, namely how should we develop grounded criteria for assessing fluency in spoken language performance. In Chapter 6 she examines the relationship between small words such as really, I mean and oh and fluency at different levels of ability. According to Hasselgreen such smallwords are present with high frequency in the spoken language and help to keep our speech flowing, although they do not necessarily impact on the content of the message itself. A major contribution of this monograph is the way she locates her argument in relevance theory as the most cohesive way of explaining how smallwords work as a system for effecting fluency by providing prototypical linguistic cues to help in the process of interpreting utterances.

In Chapter 7, based on a large corpus, she reports her research into the extent to which students taking the EVA test used smallwords. She used three groups of students: British native speaker schoolchildren of 14–15 years of age, and a more fluent and a less fluent group of Norwegian schoolchildren of the same age allocated on the basis of global grades in the speaking test. The results support the case that the more ‘smallwords’ a learner uses, the better
their perceived fluency. Critically she found that the more fluent speakers of English clearly used this body of language more frequently than high and low achieving Norwegian learners, and the range of the words they used was larger especially in turn-internal position to keep going. The more fluent learners used smallwords in a more nativelike way overall and in most turn positions than the less fluent, and also in terms of the variety of forms used and the uses to which they were put. More nativelike quantities and distribution of smallwords ‘appear to go hand in hand with more fluent speech’. The clear implication is that because small words make a significant contribution to fluent speech, such features have an obvious place when developing effective fluency scales. In Chapter 8 she analyses in more detail how students use their smallwords in helping create fluency in communication, what smallwords actually do, providing further corroboration of the findings in Chapter 7.

In Chapter 9 she looks at background variables in relation to small word use such a gender and context, and considers the acquisition of small words. She then looks at the implications of the findings of her research for language education, assessment (task and criteria) and for teaching and learning and in Chapter 10 she summarises her data in relation to the original research questions.

This volume presents the reader with a valuable framework for thinking about test validation and offers a principled methodology for how one might go about developing criteria for assessing spoken language proficiency in a systematic, empirical manner.

Cyril Weir
Michael Milanovic
Cambridge 2004