Examing Writing

Research and practice in assessing second language writing
Examing Writing

Research and practice in assessing second language writing

Stuart D Shaw
Validation Officer
University of Cambridge ESOL Examinations

and

Cyril J Weir
Powdrill Professor in English Language Acquisition
University of Bedfordshire
Contents

List of abbreviations vi
Series Editors’ note ix
Acknowledgements xiii

1 Introduction 1
2 Test-taker characteristics 17
3 Cognitive validity 34
4 Context validity 63
5 Scoring validity 143
6 Consequential validity 218
7 Criterion-related validity 229
8 Conclusions and recommendations 241

Appendices
Appendix A: Sample Writing tasks from Main Suite examinations 258
Appendix B: Candidate Information Sheet 272
Appendix C: Revised method of assessment for IELTS Writing 274
Appendix D: Current Cambridge ESOL rating procedures and marking models 275
Appendix E: Standard procedures for the production of Writing examination materials 284
Appendix F: Minimum Professional Requirements for examiners of ESOL Writing tests 289
Appendix G: Computer-based (CB) exam versions: Issues and directions 294
Appendix H: Electronic Script Management in Cambridge ESOL: Issues and directions 300

References 315
Index 339
Abbreviations

AE  Assistant Examiner
ALTE  Association of Language Testers in Europe
ANOVA  Analysis of variance
ANCOVA  Analysis of covariance
APE  Assistant Principal Examiner
AWL  Academic Word List
BEC  Business English Certificates
BMF  Batch Monitoring Form
BNC  British National Corpus
BULATS  Business Language Testing Service
CAE  Certificate in Advanced English
CB  Computer-based
CB IELTS  Computer-based International English Language Testing System
CB PET  Computer-based Preliminary English Test
CBT  Computer-based testing
CCSE  Certificates in Communicative Skills in English
CEFR  Common European Framework of Reference
CELS  Certificates in English Language Skills
CET  College English Test
CIS  Candidate Information Sheet
CLC  Cambridge Learner Corpus
CM  Clerical Marker
CMS  Clerical Marking Supervisor
Co-Ex  Co-ordinating Examiner
CPE  Certificate of Proficiency in English
CRELLA  Centre for Research in English Language Learning and Assessment
CSW  Common Scale for Writing
CUEFL  Communicative Use of English as a Foreign Language
DIF  Differential Item Functioning
EAP  English for Academic Purposes
EAQUALS  The European Association for Quality Language Services
EFL  English as a Foreign Language
ELT  English Language Teaching
Abbreviations

EM       Examinations Manager
EPS      Examinations Processing System
ERM      Electronic Return of Marks
ESL      English as a Second Language
ESLPE    English as a Second Language Placement Examination
ESM      Electronic Script Management
ESOL     English for Speakers of Other Languages
ESP      English for Specific Purposes
ETS      Educational Testing Service
FCE      First Certificate in English
FSI      Foreign Service Institute
FUEL     File Uploads from External Locations
GMAT     Graduate Management Admission Test
GMS      General Mark Scheme
IATM     Instrument for the Analysis of Textbook Materials
IEA      Intelligent Essay Assessor
IELTS    International English Language Testing System
IIS      IELTS Impact Study
ILEC     International Legal English Certificate
ILSSIEA  Instructions to Local Secretaries, Supervisors and Invigilators for Examination Administration
IRT      Item Response Theory
KET      Key English Test
LIBS     Local Item Banking System
LSA      Latent Semantic Analysis
LTRC     Language Testing Research Colloquium
MFI      Mark from Image
MFO      Mark from Object
MFR      Multi-faceted Rasch
MFRM     Multi-faceted Rasch Measurement
MFS      Mark from Script
MS       Main Suite
NLP      Natural Language Processing
NNS      Non-native speaker
NS       Native speaker
OMR      Optical Mark Reader
PA       Paper Administrator
PE       Principal Examiner
PEG      Project Essay Grader
PET      Preliminary English Test
QPP      Question Paper Production
QPT      Quick Placement Test
RCEAL    Research Centre for English and Applied Linguistics
Abbreviations

RITCME  Recruitment, Induction, Training, Co-ordination, Monitoring, Evaluation
RNIB    Royal National Institute for the Blind
RTL     Regional Team Leader
SEM     Standard Error of Measurement
SO      Subject Officer
TCT     Text Categorisation Techniques
TEEP    Test in English for Educational Purposes
TKT     Teaching Knowledge Test
TL      Team Leader
TOEFL   Test of English as a Foreign Language
TSMS    Task Specific Mark Scheme
TWE     Test of Written English
UCLES   University of Cambridge Local Examinations Syndicate
VRIP    Validity, Reliability, Impact, Practicality
YLE     Young Learners English Tests
Cambridge ESOL has long experience of the direct assessment of second language writing ability going back to the introduction of the Cambridge Proficiency in English (CPE) examination almost a century ago. In 1913 CPE required test takers to complete a two-hour English Essay, a Writing task modelled on the traditional UK school/university-based assessments of the time. By 1938 the CPE Writing component had been renamed English Composition; it included a new summary Writing task alongside the established essay and the time allocation had increased to two and a half hours. When the Lower Certificate in English (later First Certificate – FCE) was introduced in 1939 it incorporated an English Composition and Language paper lasting two hours; candidates were provided with a choice of subjects for a free composition, such as a letter or an essay on a given subject.

Since then a direct test of second language writing (and of speaking) ability has been added to subsequent examinations developed by Cambridge as and when this has been appropriate. The examination board’s commitment over many decades to direct performance assessment reflects a strong view (or construct) of proficiency as being about the ability to use language rather than simply possess knowledge about language. Individual examinations adopt an approach to assessing writing ability that is appropriate to the proficiency level, test purpose, context of use, and test-taking candidature for which they are designed; the approach shapes features such as choice of test format, task design, assessment criteria and rating descriptors. Today the Writing components in Cambridge ESOL examinations continue to be considered as useful measures of learners’ ability to communicate in written English.

The credibility of any language examination is determined by the faithfulness with which it represents a coherent understanding and articulation of the underlying abilities or construct(s) that it seeks to measure. For example, if the construct of second language writing ability is not well defined and operationalised, then it will be difficult for examination developers to support claims they wish to make about the usefulness of their writing tests. This includes claims that the tests do not suffer from factors such as construct under-representation (i.e. the test is too narrow in focus and fails to include important elements of the construct of interest) or construct irrelevant variance (i.e. the test score is prone to systematic measurement error perhaps due to factors other than the construct of interest, such as background/cultural
knowledge or unreliable scoring). Construct under-representation and construct irrelevant variance are widely regarded as the two most important threats to construct validity.

The need for clear construct definition becomes especially important when an examination developer offers writing tests at different proficiency levels (e.g. beginner, intermediate, advanced) since it presupposes a clear understanding of how the nature of second language writing ability changes across the proficiency continuum and how this can be operationalised in terms of differentiated task demands for writing tests targeted at different levels (e.g. KET, FCE, CPE).

This volume sets out to explicate the theoretical basis on which Cambridge ESOL currently tests different levels of second language writing ability across its range of test products, particularly those within its traditional Main Suite of general English examinations (KET–CPE) which span Levels A2–C2 of the Common European Framework of Reference. It does so by presenting an explicit validation framework for the testing of Writing. Building on Weir (2005), Shaw and Weir present a socio-cognitive framework which views language testing and validation within a contemporary evidence-based paradigm. They use this framework to conduct a comprehensive description and evaluation of Cambridge ESOL’s current approach to examining the skill of second language writing according to a number of dimensions or parameters.

A comprehensive model of second language proficiency remains elusive in theoretical terms; nevertheless, international language proficiency test developers such as Cambridge ESOL need to have recourse to a well-informed and coherent language proficiency model in order to operationalise it for practical assessment purposes. Such a model needs to deal satisfactorily with the twin dimensions of: (1) aspects of cognition, i.e. the language user’s or test taker’s cognitive abilities; and (2) features of the language use context, i.e. task and situation, in the testing event and beyond the test. These two dimensions constitute two of the core components within the Cambridge ESOL view of construct definition. In the specific context of practical language testing/assessment, which is where the theoretical construct must be operationalised, there exists an important third dimension: (3) the process of marking/rating/scoring itself. In other words, at the heart of any language testing activity there is a triangular relationship between three critical components:

- the test-taker’s cognitive abilities
- the context in which the task is performed, and
- the scoring process.

These three ‘internal’ dimensions of any language test – referred to in this volume as cognitive validity, context validity and scoring validity – constitute
an innovative conceptualisation of construct validity, which has sound theoretical and direct practical relevance for language testers. By maintaining a strong focus on these three components and by undertaking a careful analysis of tests in relation to these three dimensions, it becomes possible to provide theoretical, logical and empirical evidence to support validity claims and arguments about the quality and usefulness of writing tests. Having a clear and well articulated position on the underlying construct(s) can also help guide writing test revision projects and inform any future modifications.

The symbiotic relationship between the contextual parameters laid out in the task and the cognitive processing involved in task performance is stressed throughout this volume. Language testers need to give both the socio and the cognitive elements an appropriate place and emphasis within the whole, and avoid privileging one over another. The framework reminds us that language use – and also language assessment – is both a socially situated and a cognitively processed phenomenon. The twin ‘external’ dimensions of a test which are discussed in this volume – consequential validity and criterion-related validity – also reflect this understanding of the nature of language assessment from a wider perspective. The socio-cognitive framework thus seeks to marry the individual psycholinguistic perspective with the individual and group sociolinguistic perspectives. It could be argued that the socio-cognitive approach helps promote a more ‘person-oriented’ than ‘instrument-oriented’ view of the testing/assessment process than earlier models/frameworks; it implies a strong focus on the language learner or test taker, rather than the test or measurement instrument, as being at the centre of the assessment process, and it acknowledges the extent to which that assessment process is itself part of a larger social endeavour. This humanistic tradition has been a fundamental feature of the Cambridge ESOL examinations since the earliest days.

From the Cambridge ESOL perspective, the socio-cognitive framework may be the first framework which allows for serious theoretical consideration of the issues and is at the same time capable of being applied practically – hence its relevance and value to an operational language testing context. Although other frameworks (e.g. Bachman 1990) have been extremely helpful in provoking language test practitioners to think about key issues from a theoretical perspective, they have often proved difficult to operationalise in a manageable and meaningful way in the context of large-scale, international language assessment such as that undertaken by Cambridge ESOL.

In terms of the contribution it makes to research and practice in examining second language writing, the socio-cognitive framework helps to clarify, both theoretically and practically, the various constituent parts of the testing endeavour as far as ‘validity’ is concerned. The validation process presented in this volume is conceptualised in a temporal frame thereby identifying the
various types of validity evidence that need to be collected at each stage in the
test development and post implementation cycle. Within each of these, indi-
vidual criterial parameters that help distinguish between adjacent proficiency
levels have been identified and are summarised at the end of each chapter.

The framework gives us all a valuable opportunity to revisit many of our
traditional terms and concepts, to redefine them more clearly, and to grow in
our understanding. It accommodates and strengthens Cambridge ESOL’s
existing Validity, Reliability, Impact and Practicality (VRIP) approach (see
Saville in Weir and Milanovic 2003); while seeking to establish similar evi-
dence, it also attempts to reconfigure validity to show how its constituent
parts interact with one another. The results from developing and opera-
tionalising the framework in this volume with regard to testing writing ability
in the Main Suite examinations are encouraging, and evidence to date sug-
gests that where it has been applied to other Cambridge examinations/tests it
has proved useful in generating validity evidence in those cases too, e.g. in the
International Legal English Certificate, The Teaching Knowledge Test, and
BEC and BULATS (see O’Sullivan 2006). As well as showing where current
examinations are performing satisfactorily in respect of a particular validity
parameter, areas for possible improvement are highlighted, constituting a
future research agenda in Writing not only for Cambridge ESOL but poten-
tially for the wider research community.

It would be illuminating for other examination boards offering English
language tests at a variety of proficiency levels to compare their own exams in
terms of the validity parameters mapped out in this volume. In this way the
nature of language proficiency across ‘natural’ levels in terms of how it is
operationalised through examinations/tests may be more firmly grounded in
theory and thus better understood.

Michael Milanovic
Cyril Weir
Cambridge
December 2006
Acknowledgements

In bringing this volume to fruition, the authors are deeply indebted to a great many individuals. Their patience, co-operation and expert counsel has undoubtedly contributed to the success of the volume.

The first group of individuals, instrumental in terms of shaping, reading and commenting on the whole manuscript, include Professor Liz Hamp-Lyons (University of Hong Kong); Dr Sara Cushing Weigle (Georgia State University, USA); and Carole Sedgwick (Roehampton University).

The authors are very appreciative of a number of specialists in the field of language testing who provided expert input and reflection on individual chapters: Dr John Field (Reading University) for his work on cognitive validity; Dr Norbert Schmitt (University of Nottingham) and Dr Felicity O’Dell (Cambridge ESOL Testing Consultant) for their contributions to the chapter on context validity; Dr Paul Thompson (University of Reading) for his review of the scoring validity chapter; and Dr Roger Hawkey (Cambridge ESOL Testing Consultant and University of Bedfordshire) for his reviews of both the consequential validity and criterion-related validity chapters.

Sincere thanks are also due to Dr Barry O’Sullivan (Roehampton University) and Sarah Gysen (Katholieke Universiteit, Leuven) who made a number of insightful comments on various drafts of the manuscripts.

This volume could not have been completed without the additional co-operation of numerous Cambridge ESOL personnel, many of whom enabled the authors to represent fully the various voices within the organisation. We would like first to take the opportunity to acknowledge several individuals from the Assessment and Operations Group who reviewed portions of the manuscript for its comprehensibility and accuracy: Roger Johnson (Director); Anne Gutch (Assistant Director); Edward Hackett (KET/PET Subject Manager); Margaret Cooze (KET/FCE Subject Officer); Beth Weighill (formerly FCE Subject Officer); Mick Ashton (PET Subject Officer); Cris Betts (CAE Subject Officer); and Steve Murray (CPE Subject Officer).

We wish to express our special thanks to Angela ffrench in her role as FCE/CAE/CPE Subject Manager in providing an excellent liaison mechanism between the Research and Validation Group and the Assessment and Operations Group. Her insightful and patient guidance was very much appreciated.
Acknowledgements

Particular thanks goes to the Chairs of the Writing papers for the diligent and attentive reviews of text relating to their respective areas of interest: Laura Matthews; Elaine Boyd; Diana Fried-Booth; and Annette Capel.

We are also very thankful for the thoughtful and invaluable contributions made by members of the Research and Validation Group: Nick Saville (Director) for overseeing the project; Lynda Taylor (Assistant Director) especially in her editorial capacity; and to the Senior Research and Validation Co-ordinators (Neil Jones, Hanan Khalifa Louhichi and Ardeshir Geranpayeh) for their encouragement and judicious reviewing. Additional thanks goes to Anthony Green (Validation Officer), Fiona Barker (Validation Officer) and Louise Maycock (Validation Officer) for their expert guidance in selected reviews of the volume.

We would like to acknowledge the efforts of Paul Seddon (Projects Office, Cambridge ESOL) for his review of the section relating to computer-based testing and to Meredyth Rodgers (Projects Office, Cambridge ESOL) who was kind enough to critically read those portions of the text describing areas of technological assessment (particularly Electronic Script Management).

And finally we should like to recognise the contribution of Mike Milanovic (Chief Executive of Cambridge ESOL Examinations) for his encouragement throughout the entire project, and his willingness to support future research into issues raised by this study.

The publishers are grateful to the copyright holders for permission to use the copyright material reproduced in this book. Cambridge University Press for extracts and Table 5.1 from Assessing Writing by Sara Cushing Weigle, 2002. Cambridge University Press for Table 1.1 from Common European Framework of Languages: Learning, teaching, assessment by Council of Europe, 2001. Longman Pearson for Table 2H based on Measuring Second Language Performance by Tim McNamara, 1996. Palgrave Macmillan for extracts and Figure 1.1 from Language Testing and Validation by Cyril J. Weir, 2005.