Educational Testing

A Competence-Based Approach

Text for the British Psychological Society's Certificate of Competence in Educational Testing (Level A)

James Boyle and Stephen Fisher



Educational Testing

To EAB and JF

Educational Testing

A Competence-Based Approach

Text for the British Psychological Society's Certificate of Competence in Educational Testing (Level A)

James Boyle and Stephen Fisher



© 2007 by James Boyle and Stephen Fisher Educational Testing Standards are © 2007 by the British Pyschological Society

A BPS Blackwell book

BLACKWELL PUBLISHING 350 Main Street, Malden, MA 02148-5020, USA 9600 Garsington Road, Oxford OX4 2DQ, UK 550 Swanston Street, Carlton, Victoria 3053, Australia

The right of James Boyle and Stephen Fisher to be identified as the Authors of the Material in this Work has been asserted in accordance with the UK Copyright, Designs, and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs, and Patents Act 1988, without the prior permission of the publisher.

First published 2007 by The British Psychological Society and Blackwell Publishing Ltd

1 2007

Library of Congress Cataloging-in-Publication Data

Boyle, James.

Educational testing : a competence-based approach text for the British Psychological Society's certificate of competence in educational testing (Level A) / James Boyle & Stephen Fisher.

p. cm.

Includes bibliographical references and index.

ISBN-13: 978-1-4051-4659-3 (1405146591 : alk. paper)

Educational tests and measurements—Great Britain.
Teachers—Certification—Great Britain.
Fisher, Stephen.
British Psychological Society.
Title.
LB3056.G7B69 2006

371.20943—dc22

2006001705

A catalogue record for this title is available from the British Library.

Set in 10.5 on 12.5 pt Minion by SNP Best-set Typesetter Ltd, Hong Kong Printed and bound in Singapore by Markono Print Media Pte Ltd

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy, and which has been manufactured from pulp processed using acid-free and elementary chlorine-free practices. Furthermore, the publisher ensures that the text paper and cover board used have met acceptable environmental accreditation standards.

For further information on BPS Blackwell, visit our website: www.bpsblackwell.com

Contents

Preface	vii
Acknowledgements	ix
Introduction	1
Unit 1: Defining the Assessment Needs	8
Unit 2: The Basic Principles of Scaling and Standardization	35
Unit 3: The Importance of Reliability and Validity	59
Unit 4: Deciding when Psychological Tests Should or Should Not be Used as Part of an Assessment Process	72
Unit 5: Administering Tests to One or More Students/Pupils/ Learners and Dealing with Scoring Procedures	93
Unit 6: Making Appropriate Use of Test Results and Providing Accurate Written and Oral Feedback to Clients and Candidates	102
Unit 7: Maintaining Security and Confidentiality	116
Unit 8: Putting it all Together	120

vi CONTENTS

References		144
Appendix 1:	General Information Pack: Certificate of Competence in Educational Testing (Level A)	148
Appendix 2:	Guidance for Assessors: Certificate of Competence in Educational Testing (Level A)	171
Appendix 3:	The British Psychological Society Psychological Testing Centre	199
Appendix 4:	Websites of Educational Test Publishers in the UK	200
Appendix 5:	Examples of Educational Tests Available in the UK	201
Appendix 6:	Answers to Multiple Choice Items	211
Appendix 7:	Some Useful Websites	214
Appendix 8:	Extracts from the International Testing Commission Guidelines on Test Use	216
Appendix 9:	Test Evaluation Proforma	219
Index		220

Preface

Professionals working in educational settings are required to administer, score and interpret tests. They also receive reports based upon test information from among others, educational and clinical psychologists. However, there are few opportunities to learn about the range of tests used, what they measure, and their technical adequacy, outside of professional training courses, usually in psychology.

This book is written for those who wish to remedy this by undertaking training for the British Psychological Society's Certificate of Competence in Educational Testing (CCET) (Level A) (BPS, 2004a), but will be of interest to anyone who is a stakeholder in good practice in the use of tests in education. This will include Special Needs Coordinators (SENCOs), those with a responsibility for test use in educational settings, careers guidance teachers, others involved in special needs assessments and in additional support planning, and educational administrators, as well as parents.

The book provides:

- An overview section for each of the Units of the CCET, explaining the relevance of each for test use, based upon the criteria laid down in the *Guidance for Assessors* document (BPS, 2004b) for short-term and cumulative evidence.
- A section providing model answers, worked examples, case studies and self-assessments, where appropriate, linked to assessments for each element of the Units.
- Appendices with the Competences and Guidance for Assessors documents in full, together with suggestions for further reading, useful

viii PREFACE

websites, and examples of tests that would be covered by the CCET (Level A).

As assessors will vary in how they will assess competence, it is not possible to give specific examples for each Unit of every form of assessment that might be employed in a training programme (14 methods of assessment are discussed in the *Guidance for Assessors* document, and there are more that could be employed), but we will discuss the range of means by which the core competences may be assessed by the assessor, based upon the BPS *Guidance for Assessors* document (BPS, 2004b).

Acknowledgements

We are grateful to the British Psychological Society for permission to reproduce the *Competences for Educational Testing (CCET) (Level A)*, the CCET Information Pack and the *Guidance for Assessors* documents. Thanks are also due to Jacqueline Döll, Lisa Woolfson, Dave Bartram and Pat Lindley for helpful comments on the text.

Introduction

Test use in education is of more recent vintage than the use of tests for occupational selection in China over 3,000 years ago. Educational testing has its origins in the eighteenth and nineteenth centuries with the use of tests for selection for academic as opposed to vocational or technical schooling, the identification of learning difficulties and needs for additional support, and monitoring standards in schools (Oakland, 2004).

The twentieth century saw a rapid rise in the development and use of educational and psychological tests. Oakland (2004) notes that there are 5,000 standardized instruments available, although not all are in general use. Oakland and Hu (1993) reported the findings of a survey of test use in educational settings in over 44 countries, and the results revealed that the tests most commonly used by professionals working with school students were instruments measuring intelligence, personality and achievement.

Educational tests of ability, attainment and aptitude are the most widely used in the UK and the Certificate of Competence in Educational Testing (CCET) has been developed by the British Psychological Society (BPS) to meet a need for standards for the use of tests in education in the UK by psychologists and non-psychologists alike. A *General Information Pack* (BPS, 2004a) with full details of the Certificate is available from the BPS Psychological Testing Centre, and is reproduced in Appendix 1, together with *Guidance for Assessors* (BPS, 2004b), shown in Appendix 2, and information about the BPS Psychological Testing Centre in Appendix 3. The *General Information Pack* also includes the BPS's *Code of Good Practice for Educational Testing* (see BPS, 2004a, p. 15).

In the United States, standards for test use in education were developed jointly in 1985 by the American Educational Research Association, the American Psychological Society, and the National Council on Measurement in Education and were updated in 1999 (American Educational Research Association, 1999). These standards provide not only guidance in test use across a wide range of practice settings but also technical standards for those who develop and publish tests. A number of other psychological associations and societies have followed suit, including the Canadian Psychological Association (1987), the Australian Psychological Society (1997), the New Zealand Psychological Society (2002), who subsume issues relating to test use in their Code of Ethics for psychologists, and the International Test Commission (2000). However, there has been nothing comparable for the broader range of educational test users in the UK until the introduction of the CCET.

The CCET was launched by the BPS in April, 2004 with the aim of providing test users in education in the UK with an objective means of supplying evidence of a high level of competence in the selection, administration, scoring and interpretation of results from tests of ability, attainment and aptitude, that is, 'psychological' tests used in educational settings. Based upon the earlier *Level A Competences in Occupational Psychology* (BPS, 1995), the educational testing standards consist of 71 core competences arranged in seven units, as shown in Appendix 1. The standards provide guidance for those who may wish to use tests, and in particular, criteria that may be used in the assessment of an individual's competence to use tests.

All of these competences must be demonstrated to achieve the Certificate, which indicates that the holder has acquired the basic knowledge and skills required to administer and use tests in educational settings to a quality standard.

Rationale of the CCET (Level A)

The CCET standards define those things which the BPS considers a test user should know and should be able to do in order to use tests fairly, properly and in a professional manner. They cover the minimum requirements for using most standardized group and individual tests of literacy, numeracy and curriculum subjects, together with group tests of reasoning ability and measures of aspects of social competence, and provide a framework for assessments of competence leading to the award of the CCET (Level A).

However, the CCET goes further than other standards or guidelines in two ways. First, the competences are linked to a qualification, the Certificate of Competence in Educational Testing, and second, a holder of the CCET is subject to a process of quality assurance in test use to ensure high levels of practice in the workplace.

Benefits of the CCET (Level A)

The CCET (Level A) provides test users in education with the following benefits:

- a clear specification of what he/she needs to know and be able to do to use tests properly;
- an understanding of testing as one strand of the wider process of assessment;
- an awareness of the relative merits of alternative forms of assessment;
- a clear understanding of the relationship between test use, the purpose of testing and the context in which the instruments will be used;
- a set of minimum standards for training in the use of educational tests;
- a set of specific objectives that test users on training courses can match against the content of their training course; and
- a nationally recognized transferable qualification in test use in education, which is not restricted to specific tests, publishers or trainers.

While the BPS encourages test publishers and suppliers to accept the CCET as a basis for their registration of test users, it should be noted that the CCET will not allow the holder to access certain psychological instruments, sometimes referred to as 'closed' tests, for example, measures of cognitive ability such as the Wechsler Intelligence Scale for Children Fourth UK Edition (WISC-IV^{UK}) (Wechsler, 2004) or the British Ability Scales Second Edition (BAS-II) (Elliott, 1997) or some tests of achievement, such as the Wechsler Objective Reading Dimensions (WORD) (Wechsler, 1993), the Wechsler Objective Numerical Dimensions (WOND) (Wechsler, 1996), or the Wechsler Individual Achievement Test – Second UK Edition (WIAT-II^{UK}) (Wechsler, 2005). Access to such tests is restricted to those with specialist qualifications, such as educational or clinical psychologists, as the interpretation of the results requires background knowledge of psychological theory which is not covered in the CCET.

However, there is a provision for teachers to use the WIAT-II^{UK} provided that a psychologist is satisfied with the teacher's competence to

administer the test, it is ordered by the psychologist, and is used under the supervision of the psychologist in connection with the work of the psychologist in the teacher's school.

Test suppliers may require users to have specialist training in specific instruments or to hold an additional qualification in Special Needs to access some instruments. Appendix 4 provides links to the main UK educational test publishers' websites, and Appendix 5 lists examples of published tests available to users in educational settings in the domains of cognitive abilities, reading, diagnostic tests of reading, spelling, language, number, motor skills, self-esteem/coping/emotional literacy, Attention Deficit Hyperactivity Disorder (ADHD) and Autism/Asperger's Syndrome.

Short-term and Cumulative Evidence

A distinctive feature of the CCET is the requirement that assessees provide not only *short-term evidence* of competence, for example, of the kind that could be provided following initial training, but also *cumulative evidence*, which is based upon the assessee's actual test use in his/her workplace, to demonstrate that the knowledge and skills gained during training in regard to the selection and use of tests are evident and are being applied in the work setting. Accordingly, assessors require assessees to complete both short-term and cumulative assessments and to present a portfolio of evidence based upon these assessments that meets the required standard for the award of the CCET. Short-term assessments are required for all 71 competences, but cumulative assessments are only required for Units 4, 5, 6 and 7.

Obtaining the CCET

The process for obtaining the CCET is detailed in the BPS's (2004a) General Information Pack: Certificate of Competence in Educational Testing (Level A). In summary, an assessee must undertake training and then have his or her competence affirmed by an assessor, a Chartered Psychologist whose assessment methods have been verified by the BPS. Assessors have the task of ensuring that candidates provide the necessary level of evidence to demonstrate their competence. The BPS requires assessors to have their assessments verified by trained 'Verifiers' to ensure that quality standards are met. Only Chartered Educational Psychologists with Current Certificate or Statement of Competence in Educational Testing who are also

assessors can be appointed by the BPS as a Verifier, and all Verifiers receive special training to carry out their duties.

Many trainers will also be assessors, but they do not have to be. A list of assessors can be obtained from the BPS's Psychological Testing Centre website: www.psychtesting.org.uk. Before starting any training, check with the trainer that the course will lead to eligibility for the issue of the CCET, and that the assessor holds a current assessor's Certificate.

Once the assessor is satisfied that you have achieved the necessary standard of competence for all of the 71 elements, he/she will complete an Affirmation of Competence in Educational Testing (Level A) which should then be sent to the BPS together with a fee, £65 at the time of writing, which covers the costs of issuing the Certificate. The BPS has also established a Register of Competence in Psychological Testing (Educational), to which holders of the CCET may be added for an additional annual fee, £20 at the time of writing.

Any person wishing to make use of tests in education is eligible for the CCET. But note that the Certificate is not a qualification in educational psychology, but a qualification in the theoretical background and practical aspects of educational testing.

A Note for Assessors

Any Chartered Psychologist holding a CCET or a Statement of Competence in Educational Testing (available on a time-limited basis to Chartered Educational Psychologists and/or members of the Association of Educational Psychologists) who wishes to become an assessor should contact the BPS Psychological Testing Centre (see Appendix 3 for details) to obtain copies of the Assessor's Information Pack Educational Testing Level A (BPS, 2004c) and Guidance for Assessors: Certificate of Competence in Educational Testing (Level A) (BPS, 2004b), which is reproduced in Appendix 2.

Organization of this Book

The book follows the order of CCET with the 71 core competences arranged in seven Units covering psychometrics, general theories of testing and psychological attributes, test administration, reporting and feedback:

• *Unit 1: Defining the Assessment Needs:* This Unit contains competences dealing with the purpose of testing, identifying assessment needs that

can best be met by means of a test, describing the effects of environmental factors upon measures of ability and attainment, examples of corroborating information, the differences between tests of attainment, ability and aptitude, the differences between the formative and summative uses of tests, and between norm-referenced, criterion-referenced and curriculum-based measures.

- *Unit 2: The Basic Principles of Scaling and Standardization*: This Unit deals with the statistical concepts needed to use tests in educational settings.
- *Unit 3*: *The Importance of Reliability and Validity*: This Unit deals with the psychometric concepts needed to use tests.
- Unit 4: Deciding when psychological tests should or should not be used as part of an assessment process: This Unit deals with the selection of tests and issues relating to their appropriateness in specific situations. Note that the *Guidance for Assessors* (BPS, 2004b) refers to this Unit as 'Deciding whether a particular test should or should not be used as part of an assessment process'.
- Unit 5: Administering tests to one or more students/pupils/learners and dealing with scoring procedures: This is a skill-based Unit, which assesses the test user's competence in the administration and scoring of individual and group tests.
- Unit 6: Making appropriate use of test results and providing accurate written and oral feedback to clients and candidates: This Unit emphasizes the interpersonal skills required to provide oral and written feedback together with the skills required to score and interpret test results.
- *Unit 7: Maintaining security and confidentiality:* This Unit deals with issues of security and confidentiality in regard to access to, and storage of, tests and test results

The seven Units thus cover the following four areas of test use:

- *Underpinning knowledge* the basic foundations of psychometric theory and statistics required to select an appropriate test and interpret the findings: Units 1, 2 and 3.
- *Practical task skills* of planning and conducting a test session, administering the test, scoring the results, reporting the findings, and ensuring safe and secure storage of the test and record forms (if any) and also of the results: Units 5, 6 and 7.
- *Practical application of this knowledge* with particular regard to issues concerning test selection, equity and fairness: Unit 4.

• *Decision making and intervention* – recognizing the relationship between test use and issues of how best to proceed on the basis of the findings: Units 1–7.

In what follows, note that:

- 'Assessee' refers to the person whose competence is being assessed; and
- 'Client' refers to the 'commissioning agent' the person or organization for whom the test is carried out.

Unit 1: Defining the Assessment Needs

Unit 1 of the CCET (Level A) deals with a number of key issues in testing. Assessees are expected to be aware of the nature of the relationship between testing and assessment and be able to identify assessment needs that would best be dealt with by using a test (and conversely, those for which a test would not be appropriate). They are also expected to be able to justify the choice of a particular test with reference to the learner's development and the skills being assessed, and to show how the results from tests of ability and attainment can be affected by factors such as ethnicity, gender, social class and disability, as well as more direct factors in the testing environment, such as heat, noise, test anxiety and variables associated with test administration. Assessees should also be able to identify sources of information that can corroborate the information from a test, and be able to distinguish between different types of test. We shall consider each of these issues in turn.

The Relationship between Testing and Assessment

There will be few, if any, educational test users who are unfamiliar with the process of assessment, which is defined by Salvia and Ysseldyke (2004, p. 4) as: 'a process of collecting data for the purpose of making decisions about individuals and groups'.

This identifies three aspects of assessment:

• It involves a *process over time*, not just the observation of how a learner performs on one occasion.

- It is always carried out for a reason, for example to answer particular questions such as 'How do the learner's reading skills compare with those of other pupils of the same age?' or 'What are the learner's strengths and weaknesses in regard to learning new phonic skills in reading?'
- It should test one or more specific hypotheses about the learner's performance (for example, about possible factors that may precipitate or maintain a problem or explain why a problem may have occurred), which will both increase understanding and provide support for decisions as to how best to proceed.

Real-life problems are complex with many inter-related dimensions. For example, a difficulty in reading may result in problems with self-esteem and motivation, with resultant reluctance to complete classwork and homework, possible problems in peer relationships with classmates or behaviour difficulties in class, and tensions between home and school. The process of assessment has to reflect this complexity by gathering, structuring and analysing information to identify the key dimensions of the problem and understand how they interact, so that a clearer understanding both of the problem(s) and of what must be done to improve the situation can be generated.

To achieve this, information may be gathered by education professionals in collaboration with stakeholders, such as parents/carers, the students/pupils/learners themselves, and any others who may be involved, to develop a shared understanding of the nature of the problem and of the intervention required to assist decision making and achieve change. This information can be gathered in respect of an individual, or a group, for example, a class, and is a key component of the process of teaching and learning.

Such information is commonly collected in respect of concerns about academic performance, generally on account of concern about progress in the curriculum or some aspect of basic skills in language, literacy or numeracy; problems in behaviour, such as disruption, difficulties in social relationships, or withdrawn, shy behaviour; or the effects of known physical problems, for example, with vision, hearing, motor skills or mobility, or traumatic head injury.

But as Salvia and Ysseldyke (2004, pp. 23–38) and Urbina (2004, pp. 23–6) note, it is important to ensure that we gather information *from a range of sources*, (using both available information and cumulative information, such as portfolios of evidence of the learner's work) and *in different contexts* (to make the process as representative as possible) about:

• The teaching and learning environment, for example, whether the learner's teaching arrangements and the teaching techniques and

- curriculum employed are appropriate for his/her needs in regard to interest and relevance, level of difficulty, progression, presentation, and the nature of the response required; the social context of the setting and relationships with other learners; and the organization of the settings in which the learner's teaching and learning takes place.
- Factors outside the immediate educational setting, from *home and family, school and community*, which affect the learner's performance, for example, the experience he/she brings to school from home; the effects of whole-school factors, such as ethos, policies, legislation, guidelines and school development plans; and community influences upon attitudes towards learning and achievement).
- The *learner*, for example, his/her levels of confidence and independence as a learner; motivation; knowledge and use of strategies; levels of existing skills, knowledge and understanding in regard to a task or activity; learning styles; the extent to which he/she can generalize from what has been learned by reflecting upon the activity and seeing its relevance and connections to existing knowledge; and effects and influences of gender, ethnicity and social class.

There are a number of ways in which we can do this, including:

- Direct observation of the student/pupil/learner's performance or behaviour, perhaps by means of a systematic approach such as an observation schedule to investigate specific pre-defined behaviours, such as time spent 'on task', or a less systematic approach, such as a 'running record' of everything the learner does during a specified period of time. However, as well as requiring a considerable amount of time, observations can be subject to bias resulting from subjective judgements on the basis of labels (for example, such as 'gifted' or 'dyslexic'), physical attributes, other characteristics of the learner, or from expectancy effects (Salvia and Ysseldyke, 2004, pp. 27–8), where our beliefs colour what we observe, and 'we see what we expect to'. Thus, a teacher carrying out an evaluation of a behaviour management programme in his/her class may be more likely to observe negative behaviours if they have no confidence in the programme, and in contrast, more likely to observe positives if they believe that the programme has been successful.
- Existing information, such as portfolios of the learner's work, which can be compared with those of others in the class, can also be used, although interpretation will be restricted to what information is available and to the circumstances in which it was gathered.
- *Interviews* with those who know the learner well, such as former teachers or parents, also provide useful information.

- Professional judgements of those who have worked with the learner in the past or are currently working with the learner in the here and now. However, sources of bias include the possibility that key information may be forgotten, that the informant may not recognize the importance of particular information, or that informants may be selective with the information they provide.
- Finally, a test may be used.

Test use is a sub-set of the assessment process (Salvia and Ysseldyke, 2004, p. 6). But what do we mean by a 'test'? One of the leading experts in test use, Lee J. Cronbach (1984, p. 26), noted that 'There is no fully satisfactory definition for *test*'. With this caveat, let us consider some attempts at a definition:

A test is a predetermined set of questions or tasks for which predetermined types of behavioral responses are sought. Tests are particularly useful because they permit tasks and questions to be presented in exactly the same way to each person tested. (Salvia and Ysseldyke, 2004, p. 29)

A psychological test is a measurement instrument that has three defining characteristics:

- · A psychological test is a sample of behaviour.
- The sample is obtained under standardised conditions.
- There are established rules for scoring, or for obtaining quantitative (numeric) information from the behaviour sample. (Murphy and Davidshofer, 1994, p. 3)

A psychological test is a systematic procedure for obtaining samples of behaviour relevant to cognitive or affective functioning and for scoring and evaluating these samples according to standards. (Urbina, 2004, p. 1)

From these definitions, we can see that a test is a form of systematic assessment, with standardized procedures, from which numerical scores are taken. But as Urbina (2004, p. 25) notes, use of a test differs from other assessment approaches in the following ways:

- Testing is quicker and simpler than a more general assessment approach involving a range of processes, such as interviewing or observation.
- In testing, the individual test taker is the principal source of data, whereas there are other sources of information, for example, teachers, family, peers routinely involved in a more general assessment approach.

- Testing tends to compare the test taker(s) with others, whereas assessment may also consider the uniqueness of the individual or group or the context.
- Testing is highly structured, with objective procedures, whereas assessment may be less structured and more subjective, relying more on professional judgements.

But it is worth recalling that tests not only provide quantitative data, that is, the actual scores obtained, but also qualitative data in the form of observations about the test taker's performance or how the performance was achieved. As we shall see in Unit 8, such qualitative data can be important in the interpretation of the results from a test.

There are many different types of test. Rust and Golombok (2000, pp. 38–47), for example, identify the following:

- *Norm-referenced* tests (where an individual's scores are compared with those for a representative sample from a population) versus *criterion-referenced tests* (where scores are related to an external criterion, usually the completion of a task to a predefined standard).
- Knowledge-based tests (measuring ability, attainment, achievement and aptitude) versus person-based tests (measuring attributes such as personality, attitudes, moods, self-esteem, and social competence using rating scales or other self-report measures). Knowledge-based tests are tests of maximum performance, where the score is determined by the test taker's success in completing a well-defined task, that is, a measure of what the test taker can do in the test situation. Most standardized tests used in education, such as measures of ability, attainment or achievement, are of this kind. Person-based tests are tests of typical performance, where there is no 'right' or 'wrong' answer.
- Objective tests (with pre-specified scoring criteria) versus *open-ended* tests, such as essays or projective tests, where ambiguous pictures are presented and the test taker asked to describe what they see.

Salvia and Ysseldyke (2004, p. 29) identify two related dimensions that we need to take account of in making sense of test scores: *performance standards* and *informational context*. Performance standards provide the basis for comparisons between the learner and other students/pupils (a comparative standard) or between the learner's scores from two administrations of the same test (an absolute standard). But the learner's performance must also be contextualized and related to other information gathered. Good assessment – and test use – in this way utilizes *multiple sources of*