palgrave pivot

The Philosophical Limitations of Educational Assessment

Implications for Academic Selection

Ian Cantley

palgrave macmillan

The Philosophical Limitations of Educational Assessment

Ian Cantley

The Philosophical Limitations of Educational Assessment

Implications for Academic Selection

palgrave
macmillan

Ian Cantley
School of Social Sciences, Education
and Social Work
Queen's University Belfast
Belfast, UK

ISBN 978-3-031-47020-2 ISBN 978-3-031-47021-9 (eBook)
https://doi.org/10.1007/978-3-031-47021-9

Cover illustration: Melisa Hasan

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Acknowledgements

This book was written during a period of sabbatical leave in the second semester of the 2022–2023 academic year. Although I have been working on areas allied to the philosophy of educational assessment for several years, I would have been unable to write this book without the dedicated time and space created by the sabbatical, which allowed me to focus on the research underpinning the arguments presented in the book. I am therefore extremely grateful to the School of Social Sciences, Education and Social Work at Queen's University Belfast for granting me the sabbatical, and for providing the necessary resources and funding to cover my teaching and administrative duties when I was on sabbatical. I am particularly indebted to those colleagues who assisted with the PGCE Mathematics programme during the sabbatical. I would also like to thank the various individuals who read and offered constructive comments on early drafts of the manuscript. Finally, I wish to thank my family and friends for all their encouragement and support throughout the writing of the book.

About This Book

I use philosophical analysis to argue that there are tensions associated with using the results of high stakes tests to predict students' future potential. The implications of these issues for the interpretation of test scores in general are then elucidated before I consider their connotations for academic selection. After a brief overview of the history of academic selection in the United Kingdom, and a review of evidence pertaining to its consequences, I suggest that the practice of using the results of contemporary high stakes tests to make important decisions about students incurs logical and moral problems that a conscientious educator cannot ignore. The gravity of the moral transgression depends on the purpose and significance of the test and, in the case of high stakes tests used for academic selection purposes, I argue that not only can the moral wrong be highly significant, but better solutions are within reach.

Contents

About the Author

Ian Cantley is Senior Lecturer in Education at Queen's University Belfast, Northern Ireland. His research interests are in mathematics education and the mathematical and philosophical foundations of educational measurement models. He has published numerous articles in leading international journals on both the philosophy of education and mathematics education. His work is particularly concerned with the theoretical assumptions that underpin contemporary approaches to educational assessment, methods for improving students' mathematical learning experiences at school, and gender equity issues in mathematics. Ian's teaching responsibilities include contributions to the PGCE initial teacher education programme and taught masters' programmes in education, and he supervises masters' and doctoral-level dissertations on various aspects of education.

List of Abbreviations

AQE	Association for Quality Education
FSM	Free school meals
GCE	General Certificate of Education
GCSE	General Certificate of Secondary Education
IQ	Intelligence quotient
PISA	Programme for International Student Assessment
PPTC	Post-Primary Transfer Consortium
RCT	Randomised controlled trial
RDD	Regression discontinuity design
SEAG	Schools' Entrance Assessment Group
SEN	Special educational needs
SES	Socioeconomic status

List of Figures

CHAPTER 1

Introduction

Abstract In this book, I explore the controversial world of high stakes educational testing by critically evaluating some of the philosophical assumptions upon which it is based and examining the potential ethical implications of any weaknesses in contemporary approaches to high stakes testing. Initially, I critically appraise the philosophical underpinnings of educational assessment and high stakes tests in general, before focusing on the implications of my analysis for one particular use of high stakes tests: academic selection for post-primary education. Whilst I draw extensively upon evidence pertaining to Northern Ireland, and the United Kingdom more generally, my analysis is likely to be relevant to other education systems around the world. This chapter provides important contextual information that is pertinent to the material addressed in the book.

Keywords Academic selection • Educational assessment • High stakes tests • Reliability • Validity

In today's world, academic success is more important than ever. With the increasing emphasis on obtaining educational credentials and the growing competition for jobs, students are under a lot of pressure to perform well at school. Throughout their school careers, the extent of students' learning is assessed in a variety of ways, from informal monitoring of their

I. Cantley, *The Philosophical Limitations of Educational Assessment*,
https://doi.org/10.1007/978-3-031-47021-9_1

progress by teachers to timed tests taken under examination conditions. The results obtained in some of these tests have significant import in determining the students' future educational and vocational options, and such tests are commonly referred to as high stakes tests. While proponents of high stakes testing argue that these tests constitute a fair and objective way to measure academic potential, critics warn that they can exacerbate inequality and lead to a narrow approach to education that is inappropriately focused on preparation for tests and examinations. In this book, I explore the controversial world of high stakes testing by critically evaluating some of the philosophical assumptions upon which it is based and examining the potential ethical implications of any weaknesses in contemporary approaches to high stakes testing. Initially, I critically appraise the philosophical underpinnings of educational assessment and high stakes tests in general, before focusing on the implications of my analysis for one particular use of high stakes tests: academic selection for post-primary education. Whilst I draw extensively upon evidence pertaining to Northern Ireland, and the United Kingdom more generally, my analysis is likely to be relevant to other education systems around the world.

To assure their fairness, accuracy, and trustworthiness, developers of high stakes tests are expected to ensure the tests are both valid and reliable. Validity is viewed as a fundamental requirement for high stakes tests since it is associated with the degree to which the inferences made from test scores about students' capabilities are warranted. Validity may be defined as "an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores" (Messick, 1995, p. 741). It indicates the credence that can be given to inferences about students' capabilities based on test performance. Reliability, on the other hand, refers to the dependability or consistency of the results provided by a test, or the extent to which the test would yield the same or highly similar results if repeated under similar conditions. Reliability is therefore an essential aspect of validity since unreliable scores cannot support valid inferences about students' capabilities, at least at the individual student level (Cizek, 2009). However, various philosophers of education have offered critical perspectives on validity and reliability of high stakes tests, including the possible tensions between the two concepts in the context of tests of students' skills and capabilities (Davis, 1995, 1998). Considerations of validity and reliability have important implications for those who condone the use of high stakes tests to select students for

educational or vocational opportunities, and they are an important focus of my analysis in the current work.

Most education systems around the world utilise different forms of selection to allocate places to students based on their performance in such things as formal academic examinations, aptitude tests or interviews. The use of such selection mechanisms is claimed to offer equitable, impartial, and meritocratic approaches to the allocation of student places when there is competition for access to scarce educational opportunities, such as higher education courses or vocational training programmes (Kellaghan & Greaney, 2020). However, in some countries, such as Northern Ireland, primary school students are selected based on academic capability to attend different types of post-primary schools. In the case of Northern Ireland, students can take a high stakes test of their academic capability, commonly referred to as a "transfer test", towards the end of their primary school career, usually at 10 or 11 years of age. Students' performance in this test is used to determine their eligibility for admission to academically oriented grammar schools, and those students who attain the highest scores in the test can choose to attend either a grammar school or a non-selective post-primary school. In contrast, students who attain lower scores in the test, or who do not sit it, are denied admission to a grammar school, and are normally compelled to attend a non-selective post-primary school. Grammar schools are usually viewed as being synonymous with high-performance and academic success, while non-selective schools are often deemed to cater for students with less academic prowess (Brown et al., 2021). Therefore, this leads to a situation whereby post-primary students in Northern Ireland are segregated into different types of schools according to their academic capability, as measured by a high stakes test.

The current system of academic selection in Northern Ireland can be traced to the Education Act (Northern Ireland) 1947, which followed on from the 1944 Education Act in England and Wales and led to free post-primary education for all children. As in England and Wales, grammar schools had existed in Northern Ireland prior to 1947 (1944 in England and Wales), but students were admitted to them on a fee-paying basis rather than on the basis of academic capability (Gardner, 2016). However, the 1947 Education Act led to a situation where students gained access to grammar schools based on their performance in a test of their academic capability, and a system of non-selective post-primary schools was introduced to educate those who did not secure a grammar school place. A third option, for vocational education in technical schools/colleges, was

also introduced as a consequence of the 1947 Education Act, but it did not thrive and eventually disappeared (Gallagher, 2021).

Standardised testing of cognitive capabilities originated around 2200 B.C. in China, where candidates for Chinese civil service positions were given tests of their capabilities in such diverse domains as music, archery, writing and arithmetic (Miyazaki, 1981). However, in Europe and America, standardised testing of mental capabilities did not begin until the nineteenth century. In the latter part of the nineteenth century, psychologists in both Europe and America independently sought mechanisms for measuring individual differences in mental capability. The English polymath Sir Francis Galton and the American psychologist James McKeen Cattell were pioneers of the use of intelligence tests to quantify mental capabilities. Galton held the view that mental capabilities are largely inherited, and he is the founding father of eugenics, which is associated with the study of methods for improving the human race by increasing the incidence of desirable heritable characteristics, such as high levels of mental capability. Galton's enthralment with eugenics was also embraced by his disciple Sir Cyril Burt, who advocated for the widespread use of intelligence tests to classify and label schoolchildren from an early age (Chitty, 2013).

When Alfred Binet devised and published the first intelligence quotient (IQ) test in France in 1905, it was solely intended to be used to identify those students who required additional support with their learning, rather than as a mechanism for measuring and ranking all children based on their intelligence. Indeed, Binet even cautioned against using IQ as a general measure of intelligence when he claimed: "The scale, properly speaking, does not permit the measure of intelligence, because intellectual qualities are not superposable, and therefore cannot be measured as linear surfaces are measured" (cited in Gould, 1996, p. 181). Binet considered intelligence to be too complex and multi-faceted to be captured by a single number. Nevertheless, Cyril Burt worked relentlessly to ensure that his own ideas pertaining to the innateness and measurability of intelligence were incorporated into British government policy, and his work was instrumental in preparing the ground for the grammar school academic selection process that was heralded by the 1944 and 1947 Education Acts. Burt's controversial ideas pertaining to fixed capability based on inheritance have been discredited, and the policy of academic selection using tests of cognitive capabilities was gradually eroded in much of the United Kingdom during the 1960s and 1970s, with a transition to mixed-ability