The validity of critical thinking tests for predicting degree performance: A longitudinal study

Liam O'Hare *, Carol McGuinness

Queen's University Belfast, Centre for Effective Education, Graduate School of Education, Belfast BT7 1NN, United Kingdom

A R T I C L E   I N F O

A B S T R A C T

This study explored the validity of using critical thinking tests to predict final psychology degree marks over and above that already predicted by traditional admission exams (A-levels). Participants were a longitudinal sample of 109 psychology students from a university in the United Kingdom. The outcome measures were: total degree marks; and end of year marks. The predictor measures were: university admission exam results (A-levels); critical thinking test scores (skills & dispositions); and non-verbal intelligence scores. Hierarchical regressions showed A-levels significantly predicted 10% of the final degree score and the 11-item measure of 'inference skills' from the California Critical Thinking Skills Test significantly predicted an additional 6% of degree outcome variance. The findings from this study should inform decisions about the precise measurement constructs included in aptitude tests used in the higher education admission process.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Educational context

Admission to higher education in the UK is selective. The number of places available exceeds the number of applicants; so institutions, and popular courses, must use some form of admission procedure. It is widely agreed that students should gain admission to university based on valid criteria that are relevant to the educational demands of their course of study and not on background variables such as gender, ethnicity or socioeconomic status. In the UK, the most prominent criterion for admission is based on prior academic performance in the form of A-levels[1]. Over many years, there has been much public, political and media attention focused on the fairness and validity of the current practices used to admit students in the UK. Furthermore, there is growing concern among many individuals and institutions that A-levels are now not sufficiently discriminating between the higher-performing students (McManus et al., 2005). One attempt to mitigate this has been the introduction of a new A-level category, i.e., A* which is awarded to all students who achieve 90% in their exams (Ofqual, 2011). Attempts by a recent UK Minister of Education, Michael Gove, to increase the rigour of A-level through the re-introduction of a linear vs modular structure to A-levels are ongoing.

In the early 2000s, the UK Government's reaction to these concerns was to establish an advisory group, the Admissions to Higher Education Steering Group which reported on the current admissions practices and their fairness (Schwartz, 2004), as

---

* Corresponding author. Tel.: +00 44 28 90975973.
  E-mail addresses: l.ohare@qub.ac.uk (L. O'Hare), c.mcguinness@qub.ac.uk (C. McGuinness).

[1] A-levels are the examinations taken at the end of secondary schooling in England, Wales and Northern Ireland but not in Scotland which has a different system of assessment. They cover a range of academic subjects. Student performance is graded, e.g., A and B, etc. and admissions criteria are generally described in terms of AAB, BBC and so on. It is worth noting that students are offered places on their predicted grades and not their achieved grades. However, this study uses achieved A-level scores in its analysis.

well as potential future directions. One of the options suggested was to use aptitude tests to select students, although a lack of research on the validity of such tests in the UK context was noted. Since then, there has been an upsurge in use of aptitude tests for admission, particularly into high demand courses (e.g., Medicine, see Emery & Bell, 2009; Emery, Bell, & Vidal Rodeiro, 2011) and in high demand institutions (Black, 2012).

In this context, an aptitude test generally refers to a standardised psychometric measure to assess a specific cognitive ability or personal disposition which has the potential to predict future outcomes. However, there has been very little investigation into the kinds of aptitudes that might predict successful outcomes in university study. The purpose of the current research is to show how critical thinking tests might have validity as an admissions tool for higher education through their predictive validity in relation to degree outcome, their content validity through the special prominence that critical thinking has as a desirable outcome for higher education outcome and their discriminant validity to predict differences between degree classifications, for example, distinguishing first-class degrees from other degree classes.

1.2. The predictive validity of admissions criteria on degree outcome

In the United Kingdom the Higher Education Funding Council reported national statistics comparing degree outcomes of 95,000 students graduating in 2001 with different ranges of A-level points based on their three best three A-level grades (Bekhradnia & Thompson, 2002). Using probability ratios this study showed that there is variability in degree outcome even within a group of students with high A-level points. In addition, there is a substantial body of research that has examined correlations between students' A-level grades and their subsequent degree performance. In a meta-analyses of 20 studies (60 analyses) by Peers and Johnston (1994), they reported an average correlation of $r = .28$, explaining just 8% of the variance. They also reported differences between subjects, with the science students' performances being better predicted than social science and humanities' students. Chapman (1996) confirmed these differences between subject areas looking at studies from over 21 years, reporting an average correlation $r = .47$ (22% variance) for biology, and $r = .23$ (5% variance) for politics. In another systematic review of factors associated with success at medical schools, Ferguson, James, and Madeley (2002) reported prior attainment (including A-levels) had an effect size of .30 (9% variance). Specific studies on psychology undergraduates, separated by almost 40 years, reported remarkably consistent patterns. Correlations of $r = .30$ (9% of variance) was reported by Pilkington and Harrison (1967) and $r = .32$ (10% variance) was reported by Farsides and Woodfield (2003). So looking at a range of disciplinary areas, chronological time and across institutions, it appears that A-level grades predict approximately between 5% and 20% of the variance in degree outcome; even at the high end, it appears to be an imperfect process with substantial amounts of variance left unexplained.

The methods of admission into Higher Education vary greatly among global education systems but generally countries use knowledge based attainment tests or a mix of aptitude testing and attainment tests, as in the US, Sweden and Israel. For example, in the US, aptitude tests, in the form of standardised alphanumeric reasoning tests (Scholastic Aptitude Test, renamed Scholastic Assessment Test, SATs), are combined with prior educational attainment (grade point average, GPA) for admission into higher education. There are also tests equivalent to the SATs for specific purposes, e.g., the MCAT (Medical College Admission Test) for selecting medical students.

The majority of research on predicting degree attainment with aptitude tests has occurred in the US using the SATs. Burton and Ramist (2001) reviewed studies looking at predicting success in Higher Education since 1980. They reported that SAT scores and high school performance predict a range of factors including Higher Education academic performance, non-academic accomplishment, college leadership and post college earnings. They concluded that a combination measure of SATs and high school academic record was consistently the best predictor of degree outcomes. However, in a large study in the UK, the SAT was found to have no additional predictive power on Higher Education outcomes (participation and degree class) above that already predicted by GCSEs and A-levels (Kirkup, Wheater, Morrison, Durbin, & Pomati, 2010). Furthermore, even in the US, SATS have been critiqued for being narrowly focused on a limited form of reasoning. Robert Sternberg has conducted research on enhancing the predictive validity of SATS by exploring specific cognitive constructs (analytical, practical and creative skills) in the Rainbow Project (Sternberg, 2006).

Aptitude testing for admission purposes has flourished globally in the highly competitive discipline of medicine (Higgins & Sun, 2002; Kreiter, Stansfield, James, & Solow, 2003; Ferguson, James, O'Hehir, & Sanders, 2003; McManus, Smithers, Partridge, Keeling, & Fleming, 2003; McManus et al., 2005; Parry et al., 2006; Searle & McHarg, 2003). The UK alone has seen the introduction of several tests for selecting medical students including the UKCAT (TSA, 2008) and the BMAT (Emery & Bell, 2009). There is currently substantial debate in the UK around the psychometric properties of these tests (Emery & Bell, 2011; Harden, 2011; McManus, Ferguson, Wakeford, Powis, & James, 2011a,b).

The area of aptitude testing for university admission is also of interest outside medicine. For example the international exams group Cambridge Assessment now provides several aptitude tests for general university admissions to the highly competitive institutions of Cambridge, Oxford and University College London (Black, 2012). Preliminary findings have shown that these tests have good predictive validity for first year degree performance (Emery & Shannon, 2007; Harding, 2004). Previously, The Sutton Trust commissioned a substantial literature review (McDonald, Newton, & Whetton, 2001a) and pilot study (McDonald, Newton, Whetton, & Benefield, 2001b) in the area. Their work identified that the relative predictive validity of potentially useful aptitude measures would be an important research objective. Furthermore, they specifically proposed critical thinking, compared to other forms of thinking and reasoning, as an aptitude or cognitive capacity that would likely have good predictive and content validity.