



**QUEEN'S  
UNIVERSITY  
BELFAST**

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

O'Hare, L., & McGuinness, C. (2015). The validity of critical thinking tests for predicting degree performance: A longitudinal study. *International Journal of Educational Research*, 72, 162-172.  
<https://doi.org/10.1016/j.ijer.2015.06.004>

**Published in:**  
International Journal of Educational Research

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

### **Publisher rights**

© 2015, Elsevier. Licensed under the Creative Commons Attribution -NonCommercial-NoDerivs License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

### **General rights**

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

### **Open Access**

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

The Validity of Critical Thinking Tests for Predicting Degree Performance: A  
Longitudinal Study

Liam O'Hare and Carol McGuinness  
Queen's University Belfast

Dr Liam O'Hare (Corresponding Author)

Centre for Effective Education

School of Education

Queen's University Belfast

BT7 1HL

Telephone - 0044 28 90975973

[l.ohare@qub.ac.uk](mailto:l.ohare@qub.ac.uk)

Professor Carol McGuinness

Centre for Effective Education

School of Education

Queen's University Belfast

BT7 1NN

Telephone – 0044 28 90974373

[c.mcguinness@qub.ac.uk](mailto:c.mcguinness@qub.ac.uk)

## **Abstract**

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.

Keywords: critical thinking, degree performance, longitudinal, psychology degree

## **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<sup>1</sup>. 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 (Swartz, 2004), as well as potential future directions. One of the options suggested was to use aptitude

---

<sup>1</sup> 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, 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.

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 and Bell 2009; Emery, Bell and 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 and 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-analysis of 20 studies (60 analyses) by Peers and Johnson (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 et al., 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 and Sun, 2002; Kreiter, et al., 2003; Ferguson, et al., 2003; McManus, et al., 2003; McManus, et al., 2005; Parry, et al. 2006; Searle and 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 and Bell, 2009). There is currently substantial debate in the UK around the psychometric properties of these tests (Emery and Bell, 2011; Harden, 2011; McManus et al., 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 and Shannon, 2007; Harding, 2004).

Previously, The Sutton Trust commissioned a substantial literature review (McDonald, Newton and Whetton, 2001) and pilot study (McDonald, Newton, Whetton and Benefield, 2001) 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.

### **1.3 The Content Validity of Using Critical Thinking Tests for Higher Education Admissions**

While the case for using aptitude tests in university admissions has been gaining currency for some time, there has been much less debate about the *precise* focus of the aptitude that would predict success in higher education. The discussion has generally been about the relative merits of aptitude tests in general vs prior scholastic attainment (e.g., Kirkup et al., 2010; McManus et al., 2005). Instead, perhaps the question should be more about the focus of the aptitude. For example, should the aptitude test be a test of general reasoning as traditionally measured by an intelligence test, or should it be a combination of critical reading, mathematical reasoning, and writing as in the US SATS assessment, or a test that combines critical thinking, problem-solving plus numerical and spatial reasoning as in the Cambridge Assessment test of Thinking Skills or a test of analytical, practical and creative abilities as suggested by Sternberg.

In this paper the case will be advanced that critical thinking is likely to have special educational significance for success in higher education. For example, critical thinking has always been highly valued as a desirable learning outcome for higher education, both historically in UK policy documents (e.g., Dearing, 1997), in UK



quality assurance degree procedures (e.g., QAA benchmarks) and internationally in the new emphasis on 21<sup>st</sup> century skills (Voogt and Roblin, 2010). In addition, many educators have emphasised that students who are good critical thinkers will be successful in higher education (Barbanel, 1987; Chaffee, 1997; Elder and Paul, 2003; Ennis, 1996; Facione, 2011; Feldt, 1989; Higbee, 2003; Higbee and Dwinell, 1998; James, 2002; Meyers, 1987; Paul and Elder, 1996, 2003). Also, many tertiary institutions now provide explicit critical thinking courses specifically designed to improve students' critical thinking (for a comprehensive review of these programmes and their effectiveness see Abrami, et al. 2008). However, even with this general acclaim, it is not always clear what is meant by critical thinking.

Critical thinking is a concept which has its roots in philosophy as far back as Socrates, but has been particularly emphasised by the 20<sup>th</sup> century educational philosophers such as Ennis, Scriven and Fisher. More recently has been investigated by psychologists and cognitive scientists, e.g. Halpern and Kuhn. While the philosophical tradition tends to place an important emphasis on ability to challenge assumptions, to evaluate arguments and information, and to draw inferences and justifiable conclusions (e.g., Fisher 2001), the psychological tradition refers to a broader range of thinking that includes problem-solving, decision-making, hypothesis testing and so on (Halpern, 1996). Any of these forms of thinking would be considered as an important learning outcome for students in higher education. Because of the multiplicity of definitions of critical thinking, in 1990 an expert philosophers group, led by Peter Facione, came together to develop a consensus definition of critical thinking for the purposes of education and assessment in higher education, using the Delphi method. This method consists of asking a panel of experts to first submit individual definitions on a topic. These definitions are then analysed and fed back to

the panel members in a series of rounds, where the definition is revised and fine-tuned until a consensus is reached (see Gordon, 1994 and Linstone and Turoff, 1975 for descriptions of the Delphi method). These experts produced the following working consensus definition of critical thinking in their final report.

‘We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based.’

(Facione, 1990, p.2)

The most significant recommendation made by the report was that critical thinking was constructed of two main aptitudes, i.e., critical thinking skills and critical thinking dispositions. They argued that there was little point in having the competence to think critically (i.e., skills) if a person was not willing to engage in the critical discourse (i.e., dispositions). This conceptual framework underpinned the subsequent design of two tests of critical thinking – a test of critical thinking skills (i.e., ability to think critically) and a test of critical thinking dispositions (i.e., attitude towards thinking critically).

#### **1.4 A short review of Critical Thinking Aptitude Tests available to Educators**

Critical thinking tests have a relatively short history of use. The Watson Glaser Critical Thinking Appraisal, WGCTA (Watson and Glaser 1964) was the first test developed to explicitly measure critical thinking and is used primarily by occupational psychologists for the selection and promotion of candidates into management positions (Watson and Glaser, 1991) but more recently has been used in educational contexts (Macpherson and Owen; 2010). Ennis and colleagues were the first to develop critical thinking tests specifically designed for educational usage, with the Cornell Critical Thinking Test (Ennis and Millman, 1985) and the Ennis Weir Critical Thinking Essay

Test, EWCTET (Ennis and Weir 1985). More recently, the Halpern Critical Thinking Assessment (Halpern, 2010) has become available and uses both multiple choice and open-ended questions to assess critical thinking. This dual method of assessment has been advocated as it helps assess both skills and dispositional critical thinking aptitudes (Ku, 2009). The test has also shown early evidence of discriminant validity (Marin and Halpern, 2010). The predictive validity of the selection aptitude test from Cambridge Assessment<sup>2</sup> for general admissions called the Thinking Skills Assessment (TSA, 2011) has also been confirmed. The TSA is a 90 minute 50 item multiple choice assessment which aims to test Problem solving skills, including numerical and spatial reasoning and critical thinking skills, including understanding arguments and reasoning using everyday language (TSA, 2011).

Two of the most widely used measures of critical thinking emerged from the Delphi Exercise on critical thinking in 1990. Facione and colleagues designed two standardised tests of critical thinking, i.e., The California Critical Thinking Skills Test, CCTST, (Facione, Facione, Blohm, Howard and Giancarlo, 1998) and The California Critical Thinking Disposition Inventory (CCTDI) (Facione, Facione and Giancarlo, 2000) one test for each of the two major conceptual components (skills and dispositions) derived from the Delphi exercise. These tests have been chosen for the following study for three main reasons: first, they have strong theoretical underpinning through the Delphi expert consensus; second, they measure both skills and dispositions in a standardised format which aids objective statistical interpretation; and third, there is a substantial research base using these tests, which is useful for comparison purposes (see methods section for further details).

---

<sup>2</sup> Cambridge Assessment also incorporates the Oxford Cambridge examination board (OCR), which since 1999 have delivered an AS Level Critical Thinking course and a four-unit A-Level since 2005. These courses have corresponding assessments including multiple choice, short answer and essay response questions.

## **1.5 The Current Study**

Having considered the predictive validity of aptitude testing and the content validity of critical thinking tests, the following study is important because it addresses several significant research gaps. It uses a longitudinal sample of psychology students, similar to other studies with medical students (Emery and Bell, 2009), and it examines the predictive validity of critical thinking tests. Moreover, this study aims to provide further insights by identifying the additional predictive validity, on degree outcomes, over and above that already provided by A-levels, which is the real world scenario facing admissions departments in the UK higher education system. The current investigation also quantitatively explores other important aspects of validity (i.e., content and discriminant validity) in this specific educational context. Additionally, the study scrutinises specific critical thinking skills (i.e. Evaluation and Inference skills) for their differential capacity to predict performance outcomes. Lastly, the study compares critical thinking skills tests with two other aptitude tests (fluid intelligence, critical thinking dispositions) in this practical admissions scenario.

The specific research questions are:

Do the California Critical Thinking Tests have predictive validity over and above A-Levels for predicting psychology degree performance?

Do Critical Thinking Tests have content and discriminant validity for predicting higher education performance outcomes?

## **2. Method**

### **2.1 Measures Used in Studies**

Five measures of student performance were used in this study, i.e., two measures of critical thinking, the CCTST (Facione et al. 1998) and the CCTDI (Facione et. al. 2000). Two measures of academic attainment, i.e., A-levels and Degree Marks and, one measure of general intelligence Raven's Advanced Progressive Matrices Short Form, RAPM-sf (Raven, Court and Raven, 1988).

The CCTST is a multiple choice test of critical thinking skills. The 3 subscales are labelled: Analysis; Evaluation; and Inference. The CCTDI is an assessment of critical thinking dispositions completed using a six-point likert scale. It has 7 subscales, namely: Analyticity; Inquisitiveness; Open-mindedness; Systematicity; Self-confidence (related to critical thinking); and Maturity and Truth-seeking.

The two California critical thinking tests were chosen for the study because these tools have a good theoretical basis due to their partial overlap with the Delphi report - as well as their growing usage as a research tool (Abrami, 2008). They received some psychometric refinement prior to their use in this study, as a result of negative comments about construct validity by McMorris, (1995); Fawkes, O'Meara, Webber and Flage, (2005); Callahan, 1995 and Ochoa, (1995). These tests were

explored and recalibrated to improve their reliability and validity for use with students in UK settings. For reasons of clarity, the refined versions of the tests are referred to as the CCTST-UK and the CCTDI-UK.

The only alteration in the CCTST-UK was removal of the Analysis subscale from the CCTST. The reason for removing the sub-scale was extremely poor reliability (9 items with a KR-20<sup>3</sup> = .02). The reliability of the other two sub-scales Evaluation (14 items with a KR-20 = .52) and Inference (11 items, KR-20 = .40) were poor but deemed sufficient to proceed in this exploratory study<sup>4</sup>. It is likely that the low number of dichotomous items in each sub-scale is responsible for the poor reliability coefficients as there is less variability in measures with dichotomous outcomes (i.e., the answer is simply correct or incorrect). However, the combined reliability of all the items in the two subscales i.e., Evaluation and Inference together does improve reliability (25 items, KR-20 = .60).

Regarding CCTDI-UK, several items were placed into different sub-scales to improve construct validity and a number of psychometrically poor items were removed to improve reliability. The process of psychometric refinement is not fully reported in this paper (see Author, 2004 for full description of psychometric refinement details). However, the main structural changes involved removing one psychometrically problematic sub-scale 'Maturity' from the CCTDI and the remaining CCTDI-UK sub-scales had the following reliability Cronbach's  $\alpha$ , co-efficients: CT Self-Confidence

---

<sup>3</sup> KR-20(Kuder–Richardson Formula 20 - Kuder, & Richardson, 1937) is used for the estimation of test reliability of measures with dichotomous outcomes (e.g. correct or incorrect). It is similar to Cronbach's  $\alpha$ , which is the equivalent used for continuous measures.

<sup>4</sup> Ideally, Kuder–Richardson Formula 20 and Cronbach's  $\alpha$  should be above .7 to show good reliability however this is an arbitrary cut-off point and reliability is a scale indicator (i.e., 0-1) rather than a dichotomous one (i.e., good or bad). Therefore, there is value in proceeding with low reliability but the strength of findings must be tempered by the fact that the statistical analysis would have a higher degree of instrument based error.

.85; Inquisitiveness .81; Analyticity .74; Systematicity .67; Truth Seeking .63; Open-Minded .58; Maturity .36 (removed from analysis).

A-levels are UK national exams of academic attainment (generally modular format at the time of testing) taken in year 13 and 14. For the purposes of this study total A-level points were measured on a scale of 0-30. A student's best three A-levels results were included in this total. The grades were valued as follows; an A grade equalled 10 points, a B equalled 8 and so on down to an E at 2 points. Therefore, a maximum combined total of 30 was attainable. There are problems associated with altering this ordinal scale (i.e., A, B, C grades etc.) into a ratio variable (10, 8, 6 points etc.) as the ratio scale attributes equal intervals to the scores of each of the original grades, which may not be the exact case. However, this mapping process is the procedure carried out by many universities where they convert A-level grades into points for admissions and for determining equivalence between different combinations of grades and different types of qualifications. So the mapping in this paper arguably reflects real world practice and the context in which these aptitude tests would be used.

The items in the RAPM-sf feature nine separate patterns where the participant has to select the next pattern in the sequence from a selection of eight pattern choices. Research has shown this measure of intelligence is the highest correlated measure of fluid intelligence (Carroll, 1998). The version of RAPM used in this study was the short form of the test (RAPM-sf) featuring 12 items with a corresponding maximum score of 12. Some caution should be used in the interpretation of findings based on this measure as there was evidence of ceiling effects with many students scoring highly or with perfect scores in the test.

Four of the measures (CCTST, CCTDI, A-levels and RAPM-sf) provided the predictor variables for a linear regression model. These four predictor measures were

chosen so that a wide range of ability and dispositional factors, i.e., critical thinking skills, critical thinking dispositions, fluid intelligence and prior academic attainment, could be compared for their relative importance as predictors of academic performance outcomes.

The four outcome measures collected in the study were students' marks from the end of 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> year and the students' overall degree marks. There are different durations and depth of psychology course content studied in each year. The students took three modules in psychology in their 1<sup>st</sup> year and three modules from any other discipline. The second year course was composed of the core subject knowledge required for the degree to be accredited by the professional body (the British Psychological Society). The third year was more flexible in the topics offered but provided greater depth on these specific areas. The final year also included a research thesis representing 1.5 modules out of a total of six 3<sup>rd</sup> year modules. Assessment of the course was a mixture of coursework and exams. Final degree marks were weighted 40% from 2<sup>nd</sup> year marks and 60% from 3<sup>rd</sup> year marks.

## **2.2 Participants**

Over a period of three years, a total of 109 students took part in the study. However, these numbers vary in the analyses as not all students completed all tests. For example, the reported regression model required a number of predictor measures, and has lower numbers, ranging from 83-94. Of the 109 participants 92 were female and 17 were male.

All the participants in the research were single honours (N=100) or joint honours (N=9) Psychology students at Queen's University Belfast and gained Psychology degrees at that institution. The three year degree programme adheres to the British Psychological Society's criteria for Graduate Basis for Registration (GBR).



## **2.3 Procedure**

All tests used by the investigators were administered according to the guidelines outlined in the corresponding test manuals. The CCTST-UK, CCTDI-UK and RAPM-sf were administered in a lab class in the 1<sup>st</sup> week of the students' enrolling for the degree. The A-levels and various degree marks were obtained from university records. The students were briefed before and debriefed after the administration of the tests on the research purposes, and they consented to be included in the study by returning their completed tests to the experimenter. They also had the opportunity to withdraw their consent at any stage of the process.

## **3. Results**

### **3.1 Descriptive Statistics**

Table 1 describes the students who participated, providing means, standard deviations and ranges for both outcome and predictor variables. The outcome variables, i.e., degree attainment scores were typical for those of their peers both within that year group and preceding years. Mean scores in the low 60s were typical for the psychology degree in that institution. The degree outcome variables also showed wide ranges (from <40 to >70) and normal distributions within that range. There was a range of total A-level scores from 12-30 and again the scores were normally distributed. The RAPM-sf scores showed a skewed distribution towards the top end due to the high ability of the university students. The students' scores on the CCTST-UK showed normal distributions on the two subscales (i.e., Evaluation and Inference) and the combined score (i.e., Evaluation + Inference). Due to the relative difficulty of the items on the CCTST-UK, the means were around the mid-point (Total Score = 12.31 out of 25); Evaluation score = 6.29 out of 14; and Inference score = 6.02 out of 11). Again, all 6 disposition sub-scales showed normal distributions. The dispositions sub-scales

are presented sequentially in Table 1 from highest to lowest observed means of students' scores. The highest mean for the critical dispositions scales was for the Open-minded scale (4.8 on a 6-point rating scale) and the lowest mean was for the Analyticity scale (3.4 on a 6-point rating scale).

### **3.2 The validity of the California Critical Thinking Tests for predicting psychology degree marks over and above that predicted by A-Levels**

Table 2 shows the summary results of several hierarchical regressions. For reasons of parsimony, only variables that significantly correlated with degree outcome were included in the analysis. The RAPM-sf scores and CCTDI-UK scores did not significantly correlate with degree marks and thus were eliminated at this point as potential predictors of degree performance.

The first column in Table 2 'Outcome Measure' is overall average degree marks. The second column entitled 'r square for A-level' shows the percentage of the variance on this outcome predicted by A-levels. The third column shows the additional percentage of the variance that is being predicted by the 'CCTST-UK Predictors'. The predictor in the first step (or block) of all regressions was 'A-levels'. Therefore, the second column 'r square change', indicates the strength of the other predictor variables, i.e., the 'added value' that each critical thinking predictor had. In practical terms, r square change shows the added benefit these tests would have if they were used alongside A-levels for predicting degree outcomes.

There are a number of points of note in this table. A-levels were found to be a significant predictor of degree outcome ( $p < .01$ ) and accounted for 10% of the variance in average degree marks. Entry level scores on the CCTST-UK Evaluation sub-scale was not a significant predictor of degree outcome ( $p = .45$ ). Entry level scores on the CCTST-UK Inference sub-scale was a significant predictor ( $p = .01$ ) and

these scores significantly predicted 6% of the variance of degree outcome over and above A-levels) bringing the total to 16% of the variance being predicted by the combined scores. Evaluation and Inference combined scores on the CCTST-UK was not a significant predictor of degree marks but was approaching significance ( $p=.06$ ) accounting for 3% of the variance.

### **3.3 The content validity of A-levels and Critical Thinking Skills for predicting a range of psychology degree outcomes**

Table 3 compares the predictive strength of A-levels and critical thinking skills on three different psychology degree outcomes. A-levels are a consistently significant predictor of degree outcomes throughout the three years. However, A-levels decrease in their predictive power over the period of the degree from 18% in first year through to 13% in second year, and 8% of third year outcomes. The critical thinking skill of 'Evaluation' does not significantly predict any of the degree outcomes. However, it predicts a greater proportion of the variance in first and second year outcomes. The most impressive critical thinking skill predictor is 'Inference'. The 11-items of the Inference scale did not significantly predict 1<sup>st</sup> year degree outcome. However, it gradually increased in its predictive power over the three years predicting 5% of second year outcomes and 10% of third year outcomes, which was 2% more than that predicted by A-level scores (8%).

### **3.4 Discriminant validity of critical thinking tests for different degree classification**

Finally, for the purpose of illustration the data was reanalysed categorically (see Figure 1, 2 and 3) to examine the discriminant validity of the various predictors (i.e., A-levels, CCTST-UK Inference and Evaluation sub-scales) on student's final degree classification. The first point to note from the three Figures, is that all the

predictors show higher scores for 1<sup>st</sup> Class Honours students than the students receiving other degree classifications.

The differences in entry A-levels points between 1<sup>st</sup> Class Honours students ( $M = 26.46$ ,  $SD = 3.38$ ,  $N = 11$ ) and students with another degree classification ( $M = 23.73$ ,  $SD = 3.62$ ,  $N = 88$ ) are statistically significant,  $t(107) = 2.57$ ,  $p = 0.011$  (effect size  $d = .76$ ). The differences also approach significance between the Inference scores for the two groups, ( $M = 6.78$ ,  $SD = 1.76$ ,  $N = 14$ , &  $M = 5.84$ ,  $SD = 1.70$ ,  $N = 101$  respectively),  $t(114) = 1.93$ ,  $p = 0.056$  (effect size  $d = .55$ ). However, no significant difference were found with the Evaluation scores, ( $M = 6.29$ ,  $SD = 2.72$ ,  $N = 14$ , &  $M = 5.96$ ,  $SD = 2.06$ ,  $N = 101$  respectively),  $t(114) = .53$ ,  $p = 0.597$  (effect size  $d = .15$ ).

However, this analysis should be viewed as a practical illustration only and not as a robust analysis because the two groups of '1st class honours' and 'another degree classification' are unbalanced in terms of numbers with the group 'another degree classification' having a much greater sample size. Furthermore, there are no control variables in this analysis as it is just a simple t-test. For these reasons this analysis is illustrative and the main hierarchical regression analyses in 3.2 and 3.3 allow for more robust statistical inferences. However, it should be noted that both analyses show a similar pattern of results with 'inference' scale and A-levels showing the greatest discriminant validity.

#### **4. Discussion**

Before discussing the findings related to the specific research questions, a brief comment will be made about the predictive validity of A-levels in this study. A-levels predicted 10% of the variance in the overall degree marks, which is remarkably consistent with previous findings for psychology undergraduate ranging over almost 40 years (9% reported by Pilkington and Harrison, 1967, and 10% reported by Farsides

and Woodfield, 2003). Although the numbers in the current study are relatively small, the findings do not suggest that the sample is atypical for psychology undergraduates. Also, the longitudinal nature of the study, with end-of-year marks for each student across three years of the degree programme, has shown that the predictive power of A-level decreased as students progressed through the degree programme from 18% of the variance explained at the end of the first year, to 8% of the variance explained at the end of the third year. This could be merely attributed to the length of time between the two points of measurement, but it is more likely due to the changing expectations of learning outcomes for students as they progress through their studies, as will be argued below.

Turning now to the first research question; the California Critical Thinking Skills Test showed significant ‘added value’ as a predictor of higher education performance outcomes when used along-side A-levels. This study showed how the comprehensive assessment programme associated with A-levels had only marginally more predictive strength on degree outcome (10%) than the additional benefit provided by the 11- multiple choice items in the CCTST Inference sub-scale (6%). Additionally, A-levels account for less of the variance in third year degree performance outcomes (8%) than the additional benefit provided by the Inference sub-scale from the CCTST (10%). As an aptitude measure, this sub-scale has performed better than the analysis of the US SATS test in UK trials. Furthermore, these figures for predictive validity compare favourably with the use of psychometric tests in occupational psychology contexts where any additional level of significant prediction of job success is highly desirable (Goodstein and Lanyon, 1999; Roberts and Hogan, 2001).

Regarding, the second research question; the evidence suggests that measures of critical thinking skills also have good content validity. Firstly, the significant

prediction of degree outcomes provides further evidence for the close association between critical thinking and desirable outcomes for higher education. Secondly, there was surprising sensitivity for the measurement construct 'Inference'. While the predictive validity of A-levels diminished over time, the predictive power of the critical thinking associated with making inferences increased, showing the highest level of prediction at the end of third year. This pattern may reflect developmental changes produced by the content of the particular course investigated in this study. For example, the second year of the degree programme is dominated by the required core psychology content and substantial demands on statistical reasoning. In contrast, the third year demands that topics are studied in more depth, more extensive reading is required, and a research based empirical project is completed. The definition of 'Inference' used in the CCTST manual seems to be better aligned with the thinking processes that are required in third year than in the second year of the degree.

'To identify and secure elements needed to draw reasonable conclusions; to form conjectures and hypotheses, to consider relevant information and to deduce the consequences flowing from data, statements, principles, evidence, judgements, questions or other forms of representation.'

(Facione, 1998, p. 6)

In addition, although the group comparison is less statistically robust than the regression, the analysis of discriminant validity on degree classification showed that scores on the inference sub-scale discriminated well between those who go on to achieve a 1<sup>st</sup> Class Honours degree compared to those with other degree classifications.

This evidence points to an important new theoretical insight of the study in that the critical thinking skill of Inference has particularly strong validity in terms of higher education outcomes. This strength extends across predictive, content and discriminant

forms of validity. Given these strong properties further attempts should be made to conceptually explore the concept of 'Inference' and improve on the reliability and construct validity of its measurement.

The practical implications of this study are particularly relevant to higher education admissions. Firstly, the evidence supports the argument of having admissions based on a student's knowledge (from their A-levels points) as well as their skills (from aptitudes tests such as the CCTST), is better than selecting on their subject knowledge alone. This conclusion is supported by the previous evidence that these two types of assessment tap into two distinct factors of cognitive performance (Authors, 2009) and thus can predict a greater portion in higher education performance variance. In addition, critical thinking dispositions and non-verbal intelligence were not found to be significant predictors of degree performance in this study. Despite the potential ceiling effects of the RAPM-sf measure this finding still highlights the need for more research when using dispositional and general intelligence measures for selecting students into higher education. However, the evidence from a systematic review has shown the significantly higher correlation of critical thinking skills with the academic success of health professional trainees rather than critical thinking dispositions (Ross, et al., 2013).

The major limitation of this study is the low reliability of the critical thinking measures used. This may simply be due to the low numbers of items in these subscales or perhaps more serious conceptual issues around construct validity. Despite the progress that was made on investigating the psychometric properties of the CCTST and CCTDI prior to this study, the reliability and construct validity of the two tests were not good. So conclusions drawn in this paper are constrained by this fact. However, it should be recognised that real decisions about degree admissions are being

made using these imperfect measures and indicators on a daily basis. So the authors consider that this research, despite its imperfections with regard to measurement reliability, explore an important 'real world' issue.

Although references are made to undergraduate UK populations, the sample is narrow and consists of psychology students, mostly female, in one UK University. The question of how generalizable the findings are to other groups of students is still open, although extensive background information on the sample has been included to aid in comparison with other similar studies (see Table 1).

It is also worth considering that the students in this study were a pre-selected sample, i.e., they had already gained places in higher education. Psychometric theory would suggest that the predictive validity of the critical thinking tests (and the A-level scores) would have increased in strength if the sample was taken from a population with a wider distribution of abilities.

Finally, the authors suggest two general areas for future research. Firstly, there is a real need for more reliable assessment of critical thinking skills and, in particular, the skill of inference which shows the greatest predictive promise in this study. The main reason for improving reliability of measures is to permit stronger causal links to be demonstrated between pre-degree critical thinking skills and eventual degree performance outcomes rather than the more cautious conclusions that we could make in this paper. Improved reliability of critical thinking skills measures could be produced by greater conceptual refinement and more consistent measurement. For example, simply increasing the number of items in tests purporting to measure a particular critical thinking skill would be one practical step. Secondly, another area for future research is broadening the areas of measurement in terms of predictors and outcomes. The final model of critical thinking used in this study for predicting degree



outcomes is based on two critical thinking constructs from the CCTST-UK, i.e., Evaluation and Inference. Frameworks like those by Ennis (1996), Paul (1993), Kuhn (1999), Halpern (1996), Fisher and Scriven (1997) and others provide a rich source of alternative potential constructs, for example, analysis, decision making, clarity, accuracy, and communication. In addition, higher education outcomes other than degree performance could be investigated. For example, Swartz (2003) suggested a range of other potential useful higher education outcomes including post-degree societal contribution.

## **5. Conclusion**

The evidence presented in the study is relevant for research, policy and practice. There is ongoing uncertainty around the future of A-levels in relation to university admissions. Admissions managers are the gatekeepers to a valuable resource and their decisions are partially responsible for channelling who goes on to take up leadership positions in society. Therefore, they need to ensure that the people who are admitted into higher education obtain access based on reliable and valid measures of merit, and not by other economic or cultural variables. Current practice for predicting outcomes has huge scope for improvement. Research has revealed around 80% of the variance within higher education performance outcomes is still unexplained

## References

- Authors (2009). Measuring critical thinking, intelligence and academic performance in psychology undergraduates. *The Irish Journal of Psychology*, 30(2) 121-129
- Author (2004) Measuring Critical Thinking Skills and Dispositions in Undergraduate Students. *Unpublished PhD Thesis*. Author's University
- Abrami, P.C., Bernard, R.M., Borokhovski, E., Wade, A., Surkes, M.A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, 78(4), 1102-1134.
- Barbanel, J. (1987). Critical thinking for developmental students: An integrated approach. *Research and Teaching in Developmental Education*, 4(1), 7-12.
- Bekhradnia, B., & Thompson, J. (2002). *Who does best at University?* England: Higher Education Funding Council.
- Black, B. (2012). An overview of a programme of research to support the assessment of Critical Thinking. *Thinking Skills and Creativity*, 7(2), 122-133.
- British Psychological Society (2003). *Regulations and reading lists for the qualifying examination*. Leicester: British Psychological Society.
- Burton, N.W., and Ramist, L. (2001) *Predicting success in college. SAT studies of classes graduating since 1980*. New York: College Board Research Report 2001-2, College Entrance Examination Board 2001.
- Callahan, C.M. (1995). Review of the California Critical Thinking Disposition Inventory. In Conoley, J. C., and Impara, J. C. (Eds.). (1995). *The twelfth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Carroll, J. B. (1998). Human cognitive abilities. *Human cognitive abilities in theory and practice*, 5-24.

Chapman, K. (1996). Entry qualifications, degree results and value-added in UK Universities. *Oxford Review of Education*, 22(3) 251.

Chaffee, J. (1997). *Thinking Critically (5<sup>th</sup> Ed)*. Boston: Houghton Mifflin.

Dearing, R. (1997). *Higher Education in the Learning Society (The Dearing Report)*. London Report of the National Committee of Inquiry into Higher Education. London: HMSO.

Elder, L. & Paul, R. (2003). Critical thinking: teaching students how to study and learn (part 4). *Journal of Developmental Education*. 27(1) 36-37.

Emery, J.L. & Shannon, M.D. (2007). *Summary of research into the predictive validity of the TSA*. Retrieved 13<sup>th</sup> May 2014, from [http://prd.admissionstests.cambridgeassessment.org.uk/images/37863-tsa\\_predictive\\_validity\\_combined\\_analysis\\_03\\_10\\_07.pdf](http://prd.admissionstests.cambridgeassessment.org.uk/images/37863-tsa_predictive_validity_combined_analysis_03_10_07.pdf)

Emery J.L., & Bell J.F. (2009). The predictive validity of the BioMedical admissions test for pre-clinical examination performance. *Medical Education* 43,557–564.

Emery J.L., & Bell J.F. (2011). Comment on I C McManus, Eamonn Ferguson, Richard Wakeford, David Powis and David James Predictive validity of the BioMedical Admissions Test (BMAT): An Evaluation and case study. *Medical Teacher* 33,58–59.

Emery, J.L., Bell, J.F. and Vidal Rodeiro, C.L. (2011). The BioMedical Admissions Test for medical student selection: Issues of fairness and bias. *Medical Teacher*, 33, 62-71.

Ennis, R. (1996). *Critical Thinking*. Upper Sadle River New Jersey; Prentice Hall.

Ennis, R. and Millman J. (1985). *Cornell Critical Thinking Test, Level X*. Pacific Grove, California: Midwest Publications.

Ennis, R. and Weir, E. (1985). *Ennis-Weir Critical Thinking Essay Test*. Pacific Grove, California: Midwest Publications.

Facione, P. A. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations*.

Facione, P. (2011). *Critical thinking: What it is and why it counts*. Millbrae, CA: Insight Assessment, California Academic Press.

Facione, P. A., Facione, N. C. and Giancarlo, C. A. (2000). *The California Critical Thinking Disposition Inventory; Test Manual: 2000 Update*. Millbrae, CA: California Academic Press.

Facione, P.A., Facione, R.N., Blohm, S.W., Howard, K., Giancarlo, C. A. F. (1998). *California Critical Thinking Skills Test: Manual (Revised)*. Millbrae, CA: California Academic Press.

Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: the roles of personality, intelligence and application. *Personality and Individual Differences*, 34(7), 1225-1243.

Fawkes, D., O'Meara, B., Webber, D., & Flage, D. (2005). Examining the exam: A critical look at the California critical thinking skills test. *Science and Education*.14, 117-135.

Feldt, R. C. (1989). Reading comprehension and critical thinking as predictors of course performance. *Perceptual and Motor Skills*, 68(2), 642-642.

Ferguson, E., James, D. & Madeley L. (2002) Factors associated with success in medical school: systematic review of the literature. *British Medical Journal*, 324, 952–957.

Ferguson, E., James, D., O'Hehir, F. & Sanders, A. (2003). A pilot study of the roles of personality, references and personal statements in relation to performance over the 5 years of a medical degree. *British Medical Journal*, 326, 429-31.

Fisher, R. (2001). Philosophy in primary schools: fostering thinking skills and literacy. *Reading*, 35(2), 67-73.

Fisher, A. & Scriven, M. (1997). *Critical thinking: Its definition and assessment*. Norwich: Edgepress.

Goodstein, L.D. & Lanyon, R.I. (1999). Applications of personality assessment to the work place: A review. *Journal of Business and Psychology*, 13(3), 291-322.

Gordon T. J. (1994). The Delphi method. In: Glenn JC, ed. *Futures research methodology series* Washington DC: United Nations University.

Halpern, D. F. (1996). *Thought and knowledge: An introduction to critical thinking 3<sup>rd</sup> Ed*. New Jersey: Lawrence Erlbaum Associates.

Halpern, D.F. (2010). Halpern Critical Thinking Assessment: Schuhfried (Vienna Test System). Retrieved 13<sup>th</sup> May 2014, from <http://schuhfried.at>

Harden, Ronald M. (2011). If it matters, it produces controversy. *Medical Teacher*. 33(1), 4-5.

Harding, R. (2004). Thinking skills tests for university admission. Retrieved 13<sup>th</sup> May 2014, from <https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/1949/1/Harding04.pdf>

Higbee, J. L. (2003). Critical thinking and college success. *Research and Teaching in Developmental Education*. 20(1), 77-80.

Higbee, J. L. & Dwinell, P. L. (1998). The Relationship Between Student Development and the Ability to Think Critically. *Research and Teaching in Developmental Education*. 14 (2), 93-97.

Higgins, L. T. & Sun C. H. (2002). The development of psychological testing in China. *International Journal of Psychology*, 37(4), 246-254.

James, P. (2002). Ideas in practice: Fostering metaphoric thinking. *Journal of Developmental Education*. 25(3), 26-33.

Kirkup, C., Wheeler, R., Morrison, J., Durbin, B. & Pomati, M. (2010). Use of an Aptitude Test in University Entrance: a Validity Study (BIS Research Paper 26). London: BIS.

Kreiter C.D., Stansfield B., James P.A. & Solow C. (2003). A model for diversity in admissions: A review of issues and methods and an experimental approach. *Teaching and Learning in Medicine*, 15(2), 116-122.

Ku, Y. L. K. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity*, 4, 70-76.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28, 16-25.

Linstone, H. A., & Turoff, M. (Eds.). (1975). The Delphi method: Techniques and applications.

Marin, L. M., & Halpern, D. F. (2010). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gain. *Thinking Skills and Creativity*, 6, 1-13.

Macpherson, K., & Owen, C. (2010). Assessment of critical thinking ability in medical students. *Assessment & Evaluation in Higher Education*, 35(1), 41-54.

McDonald, A., Newton, P. & Whetton, C. (2001). *A pilot of aptitude testing for university entrance*. Slough: NFER.

McDonald, A., Newton, P., Whetton, C. & Benefield, P. (2001). *Aptitude testing for university entrance: a literature review*. Slough: NFER.

McManus, I.C., Smithers, E., Partridge, P., Keeling, A. & Fleming P.R. (2003). A-levels and intelligence as predictive of medical careers in UK doctors: 20-year prospective study. *British Medical Journal*, 327, 139-142.

McManus, I.C., Powis, D. A., Wakeford, R., Ferguson, E., James, D. & Richards, P. (2005). Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *British Medical Journal*, 331, 555-559.

McManus IC, Ferguson E, Wakeford R, Powis D, & James D. (2011a). Predictive validity of the BioMedical Admissions Test (BMAT): An Evaluation and Case Study. *Medical Teacher* 2011a; 33: 53–57.

McManus IC, Ferguson E, Wakeford R, Powis D, & James D. (2011b). Response to comments by Emery and Bell. *Medical Teacher* 33, 60–61.

McMorris, R. (1995). Review of the California critical thinking skills test. In Conoley, J. & Impara, J. (eds.) (1995). *The Twelfth mental measurements yearbook*. Lincoln, Nebraska: University of Nebraska Press.

McPherson, K. & Owen, C. (2010). Assessment of critical thinking ability in medical students. *Assessment and Evaluation in Higher Education*, 35(1), 41-54.

Meyers, C. (1987). *Teaching students to think critically: A guide for faculty in all disciplines*. San Francisco: Jossey Bass.



Ofqual (2011). A\* at A level FAQs. Retrieved on 13<sup>th</sup> May 2014, from <http://www.ofqual.gov.uk/help-and-support/94-articles/301-a-at-a-level-faqs>

Ochoa S.H. (1995). Review of the California critical thinking disposition inventory. In Conoley, J. C., and Impara, J. C. (Eds.) (1995). *The twelfth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Parry, J., Mathers, J., Stevens, A., Parsons, A., Lilford, R., Spurgeon, P. & Thomas, H. (2006). Admissions processes for five year medical courses at English schools: A review. *British Medical Journal*; 332, 1005-1009.

Paul, R. W. (1993). Critical thinking: what every person needs to survive in a rapidly changing world. Rohnert Park, CA: Center for Critical Thinking and Moral Critique, Sonoma State University.

Paul, R. & Elder, L. (1996). Critical thinking: Rethinking content as a mode of thinking. *Journal of Developmental Education*. 19(3), 32-33.

Paul, R. & Elder, L. (2003). Critical Thinking: Teaching Students How To Study and Learn (Part 3). *Journal of Developmental Education*. 26(3), 36-37.

Peers, I. S. & Johnston, M. (1994). Influence of learning context on the relationship between A-level attainment and final degree performance: A meta-analytic review. *The British Journal of Educational Psychology*, 64 (11) 1-18

Pilkington, G. W. & Harrison G. J. (1967). The relative value of two high level intelligence tests, advanced level, and first year university examination marks for predicting degree classification. *The British Journal of Educational Psychology*, 37(3) 382 -292

Raven, J. C., Court, J. & Raven, J. (1988). *Manual for raven's advanced progressive matrices and vocabulary scales*. London: H.K. Lewis and Co.

Roberts, B. W. & Hogan, R. (2001). *Personality psychology in the workplace*. Washington, DC: American Psychological Association:.

Ross, D., Loeffler, K., Schipper, S., Vandermeer, B., & Allan, G. M. (2013). Do scores on three commonly used measures of critical thinking correlate with academic success of health professions trainees? A systematic review and meta-analysis. *Academic Medicine*, 88(5), 724-734.

Searle, J. & McHarg, J. (2003). Selection for medical school: just pick the right students and the rest is easy! *Medical Education*, 37(5), 458-463

Sternberg, R. J., & The Rainbow Project Collaborators. (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical and creative skills. *Intelligence*, 34, 321–350.

Schwartz, S. (2004). *Fair admissions to Higher Education: Recommendations for good practice*. Admissions to Higher Education Steering Group, Chaired by Professor Steven Schwartz. Retrieved 13<sup>th</sup> May 2014, from [www.admissions-review.org.uk](http://www.admissions-review.org.uk)

Tomlinson, M. (2002) *Inquiry into A-level Standards*. DFES. Retrieved 30<sup>th</sup> May 2014, from <http://image.guardian.co.uk/sys-files/Education/documents/2002/12/03/alevelinquiry.pdf>

TSA (2008). Welcome to Cambridge Assessment Admissions Tests. Retrieved 13<sup>th</sup> May 2014, from <http://www.admissionstests.cambridgeassessment.org.uk/adt/>

Voogt, J., & Roblin, N. P. (2010). 21st century skills. Discussienota. Enschede: Universiteit Twente iov Kennisnet.

Watson, G. & Glaser, E. (1964). *Watson-Glaser Critical Thinking Appraisal Manual for Versions Ym and Zm*. New York: Harcourt, Brace and World.

Watson, G. & Glaser E. (1991). *Watson-Glaser Critical Thinking Appraisal Manual Form C*. Kent: The Psychological Corporation, Harcourt Brace Jovanovich.

**Table 1 Descriptive statistics of students' scores (mean, standard deviation, range of observed scores) on all study measures**

<b>Predictor Variables</b> <b>*Max. Score obtainable</b>	<b>Mean</b>	<b>S.D.</b>	<b>Range of Observed Scores</b>
<b>Outcome Measures</b>			
1 <sup>st</sup> year average *100 (109)	<b>59.33</b>	<b>5.48</b>	<b>46.67 – 72.00</b>
2 <sup>nd</sup> year average *100 (109)	<b>61.96</b>	<b>6.16</b>	<b>34.00 - 79.33</b>
3 <sup>rd</sup> year average *100 (100)	<b>63.33</b>	<b>6.33</b>	<b>38.83 - 80.83</b>
Degree Average *100 (109)	<b>61.40</b>	<b>6.21</b>	<b>44.00 – 78.00</b>
<b>Predictors</b>			
A-level points *30 (N=109)	<b>24.06</b>	<b>3.69</b>	<b>12 - 30</b>
RAPM-sf *12 (N=94)	<b>10.54</b>	<b>1.22</b>	<b>6 - 12</b>
CCTDI-UK (N=94)			
CCTST-UK Evaluation *14	<b>6.29</b>	<b>2.04</b>	<b>2 - 11</b>
CCTST-UK Inference * 11	<b>6.02</b>	<b>1.74</b>	<b>3 - 10</b>
CCTST-UK Evaluation Inference Combined 25*	<b>12.31</b>	<b>3.17</b>	<b>6 - 21</b>
CCTDI-UK (N=95) *6			
Open-Minded	<b>4.80</b>	<b>.68</b>	<b>3.00 - 6.00</b>
Systematicity	<b>4.20</b>	<b>.68</b>	<b>2.17 - 6.00</b>
Inquisitiveness	<b>4.09</b>	<b>.84</b>	<b>1.86 - 5.86</b>
CT Self-Confidence	<b>3.92</b>	<b>.65</b>	<b>2.00 - 5.45</b>
Truth-seeking	<b>3.84</b>	<b>.63</b>	<b>2.00 - 5.29</b>
Analyticity	<b>3.64</b>	<b>.82</b>	<b>1.29 - 5.43</b>

**\*refers to maximum score on the scale**

**Table 2 Summaries of hierarchical regressions on overall degree outcome showing  $r^2$  change and for each of the predictors when A-levels are controlled.**

\* = significant predictor at .05 level

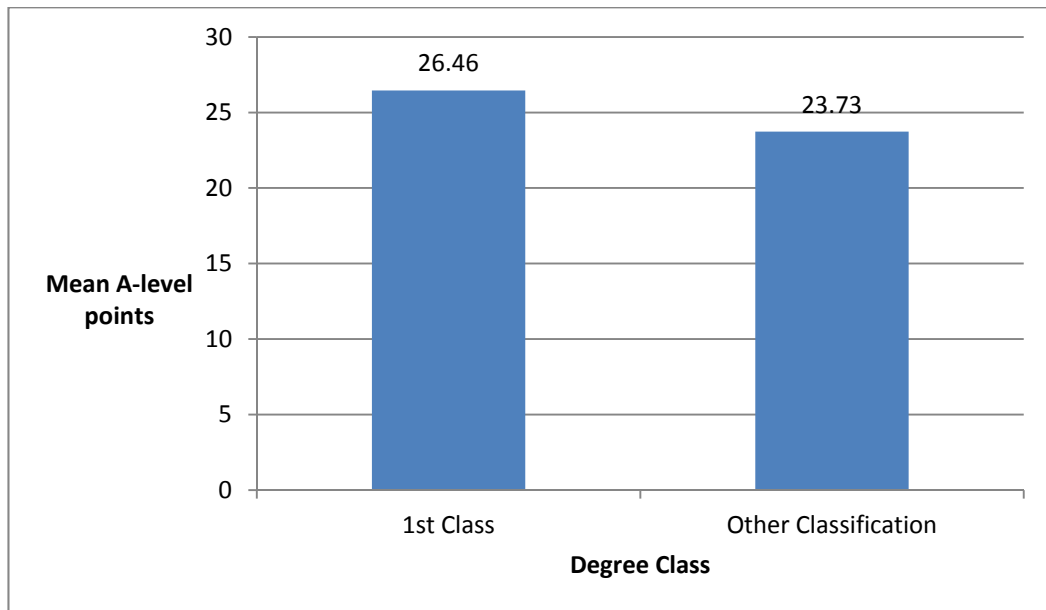
<b>Outcome Measure</b>	<b><math>r^2</math> for A-level</b>	<b>CCTST-UK Predictor</b>	<b><math>r^2</math> change (i.e., additional prediction)</b>
<b>Degree Average</b>	.10*	Entry (N=93) <b>CCTST-UK Evaluation</b>	.01
	.10*	Entry (N=93) <b>CCTST-UK Inference</b>	.06*
	.10*	Entry (N=93) <b>CCTST-UK Evaluation and Inference combined</b>	.03

**Table 3 Summaries of hierarchical regressions on three degree performance outcomes showing  $r^2$  change and for each of the predictors when A-levels are controlled.**

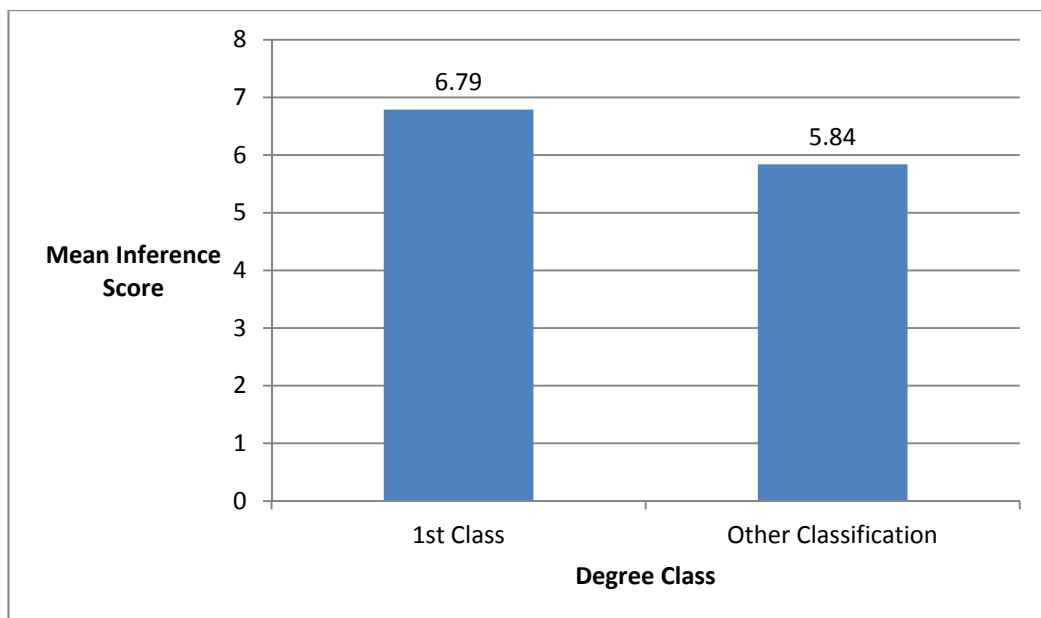
\* = significant predictor at .05 level

<b>Outcome Measure</b>	<b><math>r^2</math> for A-level</b>	<b>CCTST-UK Predictor</b>	<b><math>r^2</math> change (i.e., additional prediction)</b>
<b>Level 1 Average</b>	.18*	Entry (N=93) <b>CCTST-UK Evaluation</b>	.02
	.18*	Entry (N=93) <b>CCTST-UK Inference</b>	.01
<b>Level 2 Average</b>	.13*	Entry (N=93) <b>CCTST-UK Evaluation</b>	.03
	.13*	Entry (N=93) <b>CCTST-UK Inference</b>	.05*
<b>Level 3 Average</b>	.08*	Entry (N=84) <b>CCTST-UK Evaluation</b>	.00
	.08*	Entry (N=84) <b>CCTST-UK Inference</b>	.10*

**Figure 1 Mean A-Level Points of 1<sup>st</sup> Class Honours Compared to Other Degree Classification.**



**Figure 2 Mean CCTST-UK Inference score of 1<sup>st</sup> Class Honours Compared to Other Degree Classification.**



**Figure 3 Mean CCTST-UK Evaluation score of 1<sup>st</sup> Class Honours Compared to Other Degree Classification.**

