The Irish Journal of Psychology
2009 Vol. 30 No. 3-4 pp. 123-131

ISSN 0303-3910

Measuring critical thinking, intelligence, and academic performance in psychology undergraduates

Liam O'Hare[1] & Carol McGuinness[2]

[1]*School of Education, Queen's University Belfast*
[2]*School of Psychology, Queen's University Belfast*

The paper explores the factorial relationship between measures of critical thinking skills, non-verbal intelligence, and academic performance (A-levels and undergraduate degree marks). One hundred and twenty-nine undergraduate psychology students (94 first years and 35 third years) participated by completing two subscales of the California Critical Thinking Skills Test (CCTST) and Raven's Advanced Progressive Matrices Set 1 (APM-S1); they also provided information on their A-level points and degree marks. An exploratory factor analysis grouped the CCTST subscales of evaluation and inference with the APM-S1. The resultant factor was named 'Reasoning skills', and the A-levels and degree marks formed a second factor named 'Academic knowledge'. Furthermore, third years scored significantly higher than first years on the CCTST evaluation subscale (effect size d = 0.56), and there was a moderate effect size difference between their CCTST inference subscale scores (effect size d = 0.31) but only small effect size differences between the two groups on academic performance (d = 0.16) and APM-S1 scores (d = 0.04). It was provisionally concluded that critical thinking changes over the course of a degree and that these abilities are not well captured by traditional academic assessments. The implications of this for teaching and learning in higher education are briefly considered.

Introduction

Critical thinking in higher education: What is it and how do we assess it?

What is critical thinking? Not surprisingly, this question is at the heart of the literature on critical thinking (Beyer, 1985; Ennis, 1990; Fisher, 2001; Fisher & Scriven, 1997; Moran, 1997; van Gelder, 2001). Attempting to answer this question uncovers a wide range of issues. In its simplest form, critical thinking could be said to be - challenging a claim or an opinion (either one's own or another person's) with the purpose of finding out what to believe or do. Indeed, one of the most prolific contemporary writers on critical thinking, Robert Ennis, defined critical thinking in this way: "Critical thinking

Address correspondence to: Liam O'Hare, Centre for Effective Education, School of Education, Queen's University Belfast, BT7 1HL, Northern Ireland. Email: l.ohare@qub.ac.uk

is reasonable and reflective thinking that is focused on deciding what to believe or do." (Norris & Ennis, 1989, p. 1).

This definition captures the purpose and intention of critical thinking, but it does not say much about critical thinking as a cognitive process. However, the following definition by Halpern is more embedded in cognitive theory.

> "Critical thinking is the use of those cognitive skills or strategies that increase the probability of a desirable outcome. It is used to describe thinking that is purposeful, reasoned, and goal directed – the kind of thinking involved in solving problems, formulating inferences, calculating likelihood, and making decisions when the thinker is using skills that are thoughtful and effective for the particular context and type of thinking task" (Halpern, 1996, p. 5).

When critical thinking is considered from a psychological perspective – as a set of cognitive processes – then some typical psychological questions emerge. What kind of a construct is it? Is it one single coherent construct or are there several components involved? How can we measure these constructs? How can they be developed? Does psychology as a discipline adopt different perspectives on the nature of critical thinking? Does our education system promote it? In particular, is it important for learning and teaching in higher education? This paper attempts to elucidate some of these central questions, particularly those around the nature of the cognitive processes of critical thinking skills and how they are related to more traditional assessments of intelligence and school/university-related performance.

The notion of critical thinking raises more general questions about the nature of knowledge and reasoning. On the one hand, thinking and reasoning can be seen as a general cognitive processing ability that is readily transferable across different topics and contexts. On the other hand, thinking and reasoning can be seen as highly embedded in knowledge and disciplinary contexts such that it is only worthwhile assessing critical thinking as it relates to particular knowledge areas (e.g., psychology, history, mathematics, and art). Traditional university assessment tends to be associated with the embedded position while the psychometric measurement of critical thinking and intelligence reflects the general cognitive processing view (Carroll, 1993; Geary, 2005; Jensen, 1998).

Assessment of critical thinking at university normally occurs when an academic analyses a student's written work on a subject-specific topic for evidence of critical thinking. Difficulties encountered with this method of assessment are that students often see written work as an opportunity to show how much information s/he has acquired about a particular subject rather than as an opportunity to demonstrate critical thought. Additionally, it is often difficult for students to judge how much weight will be given to critical thinking versus knowledge in a marking scheme. Research that specifically analyses students' argument skills in essays shows that they rarely use their knowledge in an evaluative or critical manner (Anderson, Howe, Soden, Halliday, & Low, 2001). However, to help alleviate problems around critical thinking awareness and understanding in university there have been attempts to embed critical thinking more explicitly into

courses (e.g., University College Dublin in Ireland; Egan, 2007). This type of instruction is beneficial to students and educators as it gives an opportunity for discourse between the two groups about critical thinking, as well as a context for developing these important skills.

The main difference when assessing critical thinking as a general skill is that subject knowledge is no longer the focus of attention – only the critical thinking. Critical thinking tests of this kind rely on 'general knowledge' or controversial issues as prompts for critical analysis, critical argument, and critical evaluation, for example the AS/A Level GCE course in Critical Thinking delivered in some UK second level schools (OCR, 2004). Tests that measure critical thinking as a general skill can range from essay tests where the student has to construct a critical argument on a newly introduced topic, through to multiple choice tests (MCTs), where students are asked to choose from a range of options that show their ability to recognise assumptions, draw inferences, and evaluate arguments. There are also short free response type written tests for measuring general critical thinking skills, where, for example, students have to state assumptions made in a written passage or recognise conclusions reached in that passage. The grading of critical thinking essays and short free response tests is carried out in a similar way to the assessment of critical thinking within knowledge domains. In both cases, the rater has to assess the student's writing for evidence of critical thinking. However, in the former case the raters are looking for evidence of critical thinking and not evidence of subject knowledge as they would normally do when assessing students' university work. Although MCTs are more easily scored, Ennis (1996) suggests disadvantages; that is, MCTs may not be comprehensive enough to measure the full scope of critical thinking concepts. Furthermore, epistemological differences between the test producer and students taking the test can be detrimental to the student's score.

This study attempts to provide evidence for the relationships between measures of critical thinking as domain-independent multiple choice tests, and a more traditional measure of non-verbal intelligence, as well as the standard academic assessments of A-levels and degree marks. Furthermore, there is an exploration of the cross-sectional differences in these observations over the period of an undergraduate psychology degree.

The specific research questions to be addressed are: What is the factor structure underlying critical thinking, non-verbal intelligence, and academic performance? How do scores on these measures differ between a first and third year group of students?

Method

Participants and procedure

All the participants in the research were psychology students at Queen's University, Belfast. A total of 129 students took part in the study, of whom 104 (80.6%) were females and 25 (19.4%) males, and the mean age of the students was 19.71 years (SD = 2.48). All students were studying psychology and completed a battery of tests (see measures below) during psychology lectures. Ninety-four (72.8%) of the students were first year students and 35 (27.2%) were third year students.

Measures

The California Critical Thinking Skills Test, Form A (CCTST; Facione, Facione, Blohm, Howard, & Giancarlo, 1998)
The purpose of this test is to measure core critical thinking skills as outlined in the Delphi Report (Facione, 1990). The major sub-components of critical thinking suggested in the Delphi Report were Analysis, Evaluation, Inference, Interpretation, Self-regulation, and Explanation. The CCTST specifically attempts to measure Analysis, Evaluation, and Inference with three correspondingly named subscales. Facione et al. (1998) suggest the main uses of the test are as a critical thinking program evaluation tool, as an identifier of a group or individual's critical thinking strengths and weaknesses, or as a test for educational selection. However, test reviewers (McMorris, 1995; Michael, 1995) have stated concerns on the choice and placement of some of the items in this test. They suggest that in a few cases the best answer is not the one outlined in the marking scheme as the correct answer. The reviewer's (McMorris, 1995) concern is compounded by the lack of information about expert involvement on item placement in the subscales. McMorris (1995) also states that more research into the test and closer scrutiny of the items would yield a more promising test of critical thinking. A number of these points are supported in the item by item review of the CCTST Form A carried out by Fawkes, O'Meara, Weber, and Flage (2005). Nevertheless, this test was chosen because it has a good theoretical basis due to its partial overlap with the Delphi Report and its growing usage as a research tool. However, initial psychometric work with UK undergraduate samples, it was decided to use only two scales from the test, Evaluatio (10 items) and Inference (8 items). The Analysis subscale was omitted.

Advanced Progressive Matrices Set 1 (APM-S1; Raven, 1965)
This is a non-verbal test designed to assess intelligence. It has also been used as a test of inductive reasoning (Kail, 2007). Furthermore, it has been suggested that it correlates highly with 'g' (Carroll, 1993; Duncan et al., 2000). The test requires the participant to choose the most appropriate abstract design from a choice of eight to complete the series of designs. The APM-S1 is the shorter of two sets of APM available, and there are a total of 12 items which ascend in difficulty throughout. APM-S1 was chosen because of its association with non-verbal intelligence and the short version was used because it can be administered in a relatively quick time. Reduced time was important as the participants were completing a large battery of ability tests and participant fatigue was an ethical, experimental, and practical concern.

Academic performance
Two indicators of academic performance were used in the study: A-levels and university degree marks. In the UK, A-levels are the main qualifications for entry into higher education. These exams are subject-specific and often have written, oral, and coursework components. Study for these exams usually takes place over two years at the end of secondary school. The minimum requirement for entry into this psychology undergraduate degree programme at the time of testing time was three Bs or 24 points. However, some students had been admitted to the degree through access courses and had lower A-levels points.

The degree marks in this study were based on the completion of 12 modules in

psychology, six modules in year two and six in year three. These modules have a mix of exam and coursework components. The final mark is weighted 40% to year two results and 60% to year three results. Degree marks are reported as percentages, though they are not distributed across the full percentage range. The observed range for this group of students was 44-78%.

Results

The analysis is reported in three parts. The first step was to look at the distributions of the data. Table 1 shows the descriptive statistics for all the measures used in the study. The skewness statistics have been included to show how the distributions relate to the difficulty of the tests. Firstly, the APM-S1 has the greatest skewness statistic (-1.260) and it shows a skew towards the top end. Therefore, the participants found this test relatively easy and scored quite highly compared to the other tests. This is reflected in the mean of 10.56 out of a possible total of 12. However, the remaining assessments were found to be quite well distributed. Examining the skewness statistics it can be seen that the two CCTST scores show that they were skewed towards the lower end, meaning the participants found this test somewhat challenging. A-levels and degree marks were skewed towards the top, probably because of contextual factors. The majority of students were selected based on A-levels points being greater than three Bs which is towards the top end, and degree results are marked using a system which see the majority of candidates receiving 2:1 degree scores (60+), again skewing the data to the top end.

Table 1. Descriptive statistics for the five assessments used in the study.

Measure	Minimum	Maximum	Mean	Std. Deviation	Skewness
APM-S1	5	12	10.56	1.32	-1.260
CCTST (Evaluation)	1	10	5.04	1.83	0.214
CCTST (Inference)	0	8	3.71	1.58	0.244
A-level points	12	30	23.97	3.60	-0.140
Degree mark	44	78	61.33	5.94	-0.197

The next goal was to see if these measures could be reduced to a number of higher order factors. Firstly, the choice of exploratory factor analysis was considered. As at least one measure was relatively skewed (APM-S1), Principal Axis Factoring was chosen because of its consideration of non-normal distributions (Osborne & Costello, 2005). This was carried out in conjunction with a direct oblimin rotation because theoretically any higher order factors extracted should have some correlation due to all measures being a form of cognitive assessment. The scree plot slope and eigenvalues suggested that two factors should be extracted from the data, and the results can be seen in Table 2. There were no cross loadings above 0.4 on either factor. The factor correlation matrix suggested that the two factors were moderately correlated (0.3). Next the factors were named ‘Reasoning skills’ and ‘Academic knowledge’ based on the top loading measures within each factor. The rationale and implications of this are explored in more detail in the Discussion.

Table 2. Principal Axis Factor structure matrix with oblimin rotation and Kaiser normalisation.

	Reasoning Skills	Academic Knowledge
CCTST (Inference)	0.666	
APM-S1	0.493	
CCTST (Evaluation)		0.484
A-level points		0.895
Degree mark		0.416

The last part of the analysis was to conduct independent t-tests on all of the measures. The two independent groups in this instance were whether the test battery was completed on entry to the psychology degree (year one) or at the end of the psychology degree (year three).

Table 3. Means and independent samples t-test of all measures in study.

Measure	Mean 1st Year (N = 94)	Mean 3rd Year (N = 35)	t (df = 127)	Effect size (Cohen's d)
APM-S1	10.54	10.60	-0.22	0.04
CCTST (Evaluation)	4.77	5.77	-2.85*	0.56
CCTST (Inference)	3.57	4.06	-1.55	0.31
A-level points	24.06	23.71	0.49	-0.09
Degree marks	61.08	62.00	-0.78	0.16

* significant at $p < 0.05$ level

Table 3 shows that only the CCTST evaluation score was significantly different between the two groups, and the only other moderate effect size difference between the two groups was the CCTST Inference score. The implications of this are outlined in the Discussion.

Discussion

The factor analysis produced two distinct factors that were named 'Reasoning skills' and 'Academic knowledge' – although they were moderately correlated, and thus share some common variance, they still retained general independence. The psychometric tests that evaluated the more decontextualised forms of thinking and reasoning which had been presented in multiple choice test formats loaded on the first factor. The Inference subscale on the CCTST had the highest loading and defined the factor. The students' performance on this subscale was marginally poorer than their scores on the Evaluation subscale, indicating that the items were more difficult. As the names of the subscales suggest, 'inference' does require a more elaborate form of critical reasoning beyond 'evaluation'. Facione's (1990) own description of the scales (and examining the items themselves) suggests that the Evaluation subscale calls on the ability of the student to assess the credibility of statements and/or their logical strength, whilst the Inference subscale asks

the students to go one step further and to generate consequences and conclusions. Not surprisingly, the measure of non-verbal reasoning also loaded on this factor, indicating that the construct is similar to a general ability factor (see Carroll, 1993; Cooper, 1998 for more information), with CCTST measuring verbal reasoning or crystallised intelligence and the APM-S1 measuring non-verbal reasoning or fluid intelligence. Although, both the non-verbal and verbal tests loaded on the same factor, they were at a different level of difficulty for the student cohort being studied. On the non-verbal reasoning tests, the students' scores were generally high while there was clearly room for improvement in their critical thinking. Also, the moderate effect sizes in Table 3 between the CCTST scores for the first year students and the third year students suggests that the experience of higher education did exert a positive influence on the quality of this mode of thinking, but because of the cross-sectional nature of this data set, this conclusion is a cautious one, yet the two cohorts do seem well matched on the other indicators of performance.

A-level points and degree marks loaded on the second factor, 'Academic knowledge', with the factor being defined by the highest loading which was A-level performance. Both these assessments include extended writing formats, as well as a range of other assessment types (including some multiple choice elements). A-level points are the combination of the students' performance over three school subjects and, as such, probably assess the students' more generalised approach to dealing with academic forms of knowledge, while psychology degree marks are an indicator of how well the students have performed according to the marking criteria in a more focused area of study and at a level appropriate to higher education rather than secondary education. Also, evidence of critical thinking is explicitly included in the marking criteria for awarding degree classes in psychology, so it is perhaps not surprising that degree marks has a lower loading on this factor compared to A-levels. However, the modest correlation between the two factors, and the absence of any substantial cross-factor loading, do indicate that traditional forms of assessment both at secondary and higher education fail to capture sufficiently the forms of critical thinking which are considered the hallmark of achievement in higher education.

What are the implications of these findings for A-levels (or Leaving Certificate performance in the Republic of Ireland) and for higher education teaching and assessments? In the UK, A-levels are the standard method for selecting students into higher education, and more than 90% of students with two A-levels proceed to higher education (Smithers & Robinson, 1995). Over the past few years, there has been much public, political, and media attention focused on the fairness and validity of the current practices used to select students. Furthermore, there was some concern that A-levels were not sufficiently predicting future degree performance of students (Peers & Johnston, 1994; Sear, 1983). Subsequently, a variety of research and development projects have been launched to evaluate the potential for combining A-levels with some form of aptitude testing for selection. For example, the National Foundation for Educational Research (NFER) in the UK are conducting a predictive validity study, comparing the aptitude test which is used in the US, SATS (Scholastic Aptitude Test, now called Scholastic Assessment Test), with A-levels to predict degree success (Kirkup, Schagen, Wheater, Morrison, & Whetton, 2007). Also, there is much debate about selection in medical education which has resulted in the production and promotion of the UK Clinical Aptitude Test in UK

medical and dental schools (UKCAT, 2008). In both cases, the content of the test seems to be focused on forms on verbal, spatial, and quantitative reasoning that are typical of general ability tests, rather than a more precise focus on forms of reasoning that are more likely to be crucial for success in higher education and beyond, such as critical reasoning and problem solving.

In terms of teaching and assessment in higher education, there is substantial reference to the importance of critical thinking in degree benchmarking statements, learning outcomes for modules, degrees programmes, and so on. However, the extent to which direct reference to forms of critical thinking, or teaching critical thinking more explicitly, is part of the regular educational discourse in higher education lectures and tutorials is much less clear, one exception to this being the work in University College Dublin mentioned earlier (Egan, 2007).

Although higher education discourses appear to value critical thinking highly, more attention needs to be paid to how these forms of thinking are explicitly developed, how they are integrated into ongoing disciplinary-based degree programmes, and how assessments and marking schemes are designed to ensure that they are given sufficient weight in awarding degrees.

References

Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science, 29*, 1-32.

Beyer, B.K. (1985). Critical thinking: What is it? *Social Education*, April, 270-276.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor- analytic studies.* Cambridge: Cambridge University Press.

Cooper, C. (1998). *Individual differences*. London: Arnold.

Duncan, J., Seitz, R.J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F.N., & Emslie, H. (2000). A neural basis for general intelligence. *Science, 289*, 457-460.

Egan, A. (2007). *Explicit development of critical thinking in undergraduate education.* Unpublished PhD Thesis. Dublin, Ireland: University College Dublin.

Ennis, R. (1990). The rationality of rationality: Why think critically? In R. Page (Ed.), *Philosophy of education 1989* (pp. 402-405). Bloomington, Ill: Philosophy of Education Society.

Ennis, R.H. (1996). *Critical thinking*. Upper Saddle River, NJ: Prentice-Hall.

Facione, P. (1990). *The Delphi Report: Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. 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.

Fawkes, D., O'Meara, B., Weber, D., & Flage, D. (2005). Examining the exam: A critical look at The California Critical Thinking Skills Test. *Science and Education, 14*, 117-135.

Fisher, A. (2001). *Critical thinking: An introduction*. Cambridge: Cambridge University

Press.
Fisher, A., & Scriven, M. (1997). *Critical thinking: Its definition and assessment.* Norwich, UK: Edgepress.
Geary, D.C. (2005). *The origin of mind: Evolution of brain cognition and general intelligence*. Washington, DC: American Psychological Association.
Halpern, D.F. (1996). *Thought and knowledge: An introduction to critical thinking* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
Kail, R.V. (2007). Longitudinal evidence that increases in processing speed and working memory enhance children's reasoning. *Psychological Science, 18*, 312-313.
Kirkup, C., Schagen, I., Wheater, R., Morrison, J., & Whetton, C. (2007). *Use of an aptitude test in university entrance - a validity study: Relationships between SAT scores, attainment measures and background variables.* (DfES Research Report 846). London: DfES.
McMorris, R. (1995). Review of the California Critical Thinking Skills Test. In J. Conoley & J. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 143-145). Lincoln, NE: University of Nebraska Press.
Michael, W. (1995). Review of the California Critical Thinking Skills Test. In J. Conoley & J. Impara (Eds.), *The twelfth mental measurements yearbook* (pp. 145-146). Lincoln, NE: University of Nebraska Press.
Moran, A.P. (1997). *Managing your own learning at university: A practical guide.* Dublin: Betaprint.
Norris, S.P., & Ennis, R. (1989). *Evaluating critical thinking.* Pacific Grove, CA: Midwest Publications.
OCR (2004). Specifications for the OCR Advanced Subsidiary GCE in *Critical Thinking (H050) & OCR Advanced GCE in Critical Thinking (H450)*. Retrieved November 25, 2008, from http://www.ocr.org.uk/qualifications/as_alevelgce/critical_thinking/
Osborne, A.B., & Costello, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10*. Retrieved January 7, 2009, from http://pareonline.net/getvn.asp?v=10&n=7
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. *British Journal of Educational Psychology, 64*, 1-18.
Raven, J. (1965). *Progressive matrices.* London: H.K. Lewis.
Sear, K. (1983). The correlation between A-level grades and degree results in England and Wales. *Higher Education, 12*, 609-619.
Smithers, A., & Robinson, P. (1995). *Post-18 education: Growth, change, prospect.* London: The Council for Industry and Higher Education.
UKCAT (2008). *UK Clinical Aptitude Test: 2006 Annual Report*. UKCAT Board. Retrieved July 7, 2008, from http://www.ukcat.ac.uk/pdf/UKCAT%20Annual%20Report%202006%20v2.pdf
van Gelder, T.J. (2001). The reason! Project. *The Skeptic, 21*, 9-12.