Thinking Skills and Creativity 7 (2012) 122–133

Contents lists available at SciVerse ScienceDirect

Thinking Skills and Creativity

journal homepage: http://www.elsevier.com/locate/tsc

An overview of a programme of research to support the assessment of Critical Thinking

Beth Black*

Cambridge Assessment, University of Cambridge, UK

ARTICLE INFO

Article history:
Received 12 March 2012
Received in revised form 29 March 2012
Accepted 2 April 2012
Available online 16 April 2012

Keywords:
Critical Thinking assessment
Definition
Taxonomy
Impact
Validation

ABSTRACT

Cambridge Assessment has more than 20 years experience in assessing Critical Thinking (CT) in a number of diverse tests and qualifications, unrivalled by any other body within the UK. In recent years, a number of research activities have been carried out in order to support these assessments, with a focus on the validity of measurement. This paper will describe the programme of work, related to four areas of research: (1) to derive a definition, taxonomy and glossary of CT and the importance of such work not just for test design but also for syllabus design and question setting; (2) to survey the practical issues surrounding teaching CT in schools, the experiences and perceptions of teachers; (3) to examine the impact of taking A/AS levels in CT on students' performance in their other A-level subjects; and (4) to examine the predictive validity of HE admissions tests containing CT on subsequent university exams and degree performance. Finally, the paper will discuss how these research activities contribute to the validity of CT tests and qualifications.

1. Introduction

While the programme of research described here is quite diverse in terms of method and scope, an overarching purpose of all of the research around Critical Thinking assessments is validity. Here, validity is used in an educational measurement context and refers to "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests." (AERA/APA/NCME, 1999, p. 9).

Cambridge Assessment, the University of Cambridge's international examinations group, has a range of qualifications and tests which assess Critical Thinking (CT) and which have been developed over twenty years. Many of these share a common ancestor – though not all – and many have been developed by a number of common experts in the area. During this time period, the nature and purpose of these assessments have proliferated, at the same time as Critical Thinking has become a much more widely embraced educational concept.

The catalyst for much of this programme of research was a growing realisation that there was a need to take stock and, in particular, determine the degree to which there was coherence across the tests and qualifications offered across Cambridge Assessment. Cambridge Assessment has three examining bodies: (i) OCR – Oxford Cambridge and RSA – delivers public examinations in the UK including GCSEs[1] and Advanced Level qualifications (A levels)[2]; (ii) Cambridge International

* Correspondence address: Cambridge Assessment, 1 Regent Street, Cambridge CB2 1GG, UK.
E-mail address: Black.B@cambridgeassessment.org.uk

[1] GCSE – General Certification of Secondary Education – are designed to be taken by 16 year olds after a two-year course of study. These are available in a wide range of subjects.

[2] A levels (Advanced Level General Certificate of Education) – are the most popular post-16 academic route in England, Wales and Northern Ireland. They typically require two years of study beyond GCSE, with the first year of work being assessed at Advanced Subsidiary (AS) level. For further description see beginning of Section 2.

1871-1871/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.tsc.2012.04.003

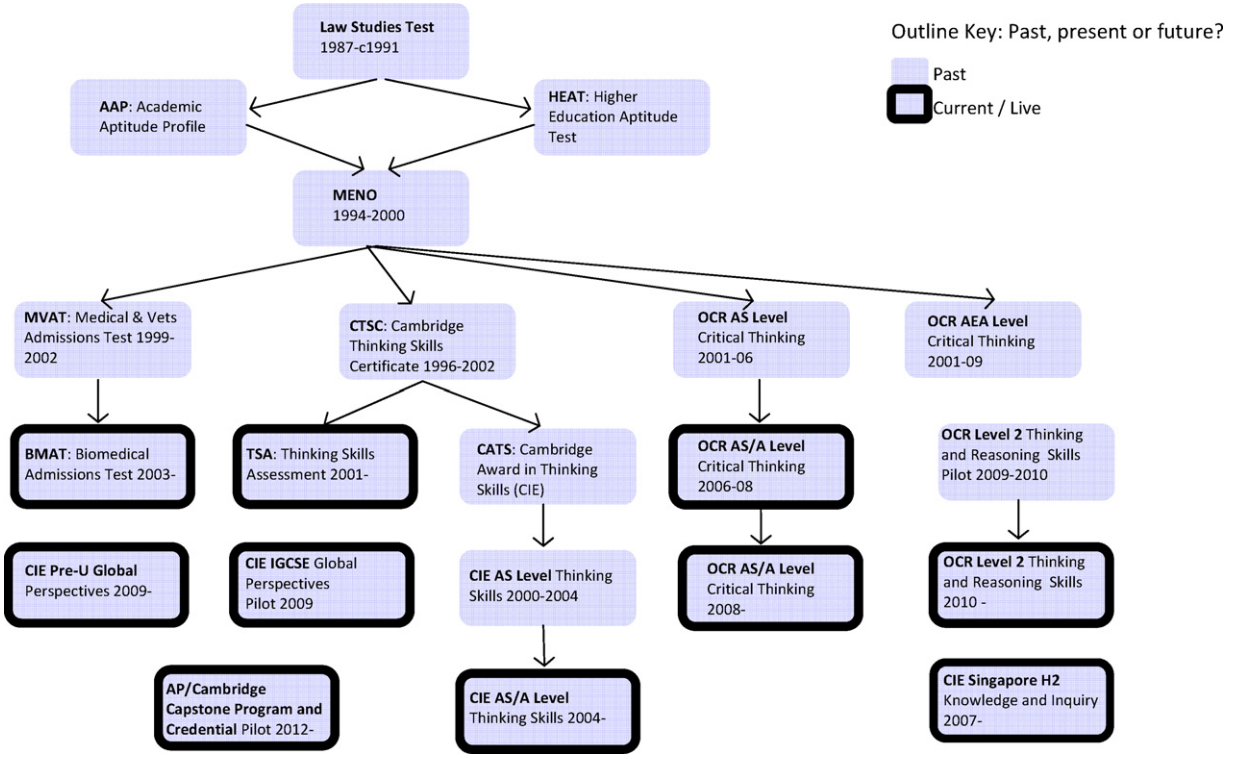

Fig. 1. A family tree of Cambridge Assessment's qualifications and tests which include Critical Thinking.

Examinations (CIE), offering a broad range of internationally recognised academic and vocational qualifications including IGCSEs[3] and International A levels taken in over 150 countries worldwide; (iii) Cambridge ESOL (English for Speakers of Other Languages), offering examinations in English language for non-native speakers of English and teaching qualifications. Fig. 1 provides a diagrammatic representation of the range of assessments and qualifications containing Critical Thinking and how they have evolved. Currently, Cambridge Assessment has four, long term, extant products (see Fig. 1): BMAT,[4] TSA,[5] CIE Thinking Skills AS/A level and OCR AS/A Level Critical Thinking, all of which share a common ancestor, namely MENO.[6] However, each assessment has a slightly different evolutionary history, tests differing aspects and subsets of Critical Thinking, and is used for different purposes and candidate types.

It was considered important that, across Cambridge Assessment's extant Critical Thinking offerings, there should be a coherent, shared understanding of the usage of the term Critical Thinking and the construct being measured. It was also important that this should also be true of any assessment of the subject developed in the future. Hence, the first research activity summarised in this article is about deriving a Cambridge Assessment definition and taxonomy of Critical Thinking, i.e., defining the overall construct. An extension to this work, also reported here, involved taking the definitions further by creating a glossary of specific terms within Critical Thinking.

The second research activity described in this paper looks at practices in UK schools and colleges in terms of delivering Critical Thinking AS/A level. This qualification saw a substantial increase in numbers particularly between 2001 and 2008. Using a large scale survey, it was investigated how schools and colleges were embracing the discipline of Critical Thinking, how teachers were themselves becoming competent in teaching the discipline and how they were resourcing the course in general for their students.

The third research activity summarised in this paper looks at the impact of taking Critical Thinking on other A level subjects. Advocates of Critical Thinking often claim this is a transferable discipline – indeed, its very raison d'être – in that CT skills can and should be transferred for use in other academic disciplines and other contexts. These transferable skills should enhance performance or create some advantage in other contexts.

The fourth area of work summarised in this article relates specifically to BMAT and TSA. These are admissions tests for Higher Education developed and administered by Cambridge Assessment. The sorts of courses and universities which use

[3] International General Certificate of Secondary Education (IGCSE) – these qualifications are comparable to UK GCSE.
[4] BioMedical Admissions Test – used as a selection test for a number of Medical and Veterinary university courses.
[5] Thinking Skills Assessment – used as a selection test for a number of highly competitive university courses.
[6] MENO is not an acronym – this test was named after one of Plato's students.

TSA and BMAT as part of their admissions process can generally be described as highly competitive (e.g., Engineering at the University of Cambridge, Biomedical Sciences at the University of Oxford). The purpose of the test as a selection instrument closely relates to its ability to predict future performance on such courses.

While this paper describes four very different activities around Critical Thinking assessments, the scope of the overall programme of work provides us with confidence around the design and purpose of the assessments and that all these activities contribute to their validity.

2. Defining Critical Thinking at Cambridge Assessment – a definition, taxonomy and glossary

> "*Are some outcomes of education too intangible to be measured? No doubt, there are some that we speak of often, like critical thinking. . ., that [is] so difficult to define satisfactorily that we have given up trying to define [it] specifically. To this extent, they are intangible [and] hard to measure.*" (Ebel, 1965)

Ebel (1965) makes a point which seems quite sensible, though pessimistic. And yet, a few decades on, the testing of Critical Thinking has become a flourishing area. As depicted in Fig. 1, during the two decades of Cambridge Assessment producing Critical Thinking assessments, the range and nature of these tests and qualifications have grown and evolved. This proliferation of tests and qualifications was despite there being no single agreed definition – there were many conflicting definitions (e.g., Dewey, 1909; Ennis, 1996; Facione, 1990; Fisher & Scriven, 1997; Glaser, 1941; Paul, 1992; Sternberg, 1986). It is perhaps fair to say that, in the absence of a single agreed definition in the area, the conception of what these tests measured had been largely transmitted implicitly through the coincidence of a core group of common experts and personnel working on these tests and writing items for them.

2.1. A definition and taxonomy of Critical Thinking

However, in order to inspect the construct validity, representativeness and coherence across such assessments, we needed an explicit working conception of the domain of CT. The main aim of this research was to create a definition and taxonomy of Critical Thinking in order to support validity arguments about Critical Thinking tests and exams. Additionally, it was also considered desirable to have the domain more clearly defined and mapped in order to guide the development of any future tests and qualifications.

More details of the method and outcomes can be found in Black (2008); but in brief, an expert panel was convened, consisting of five people with considerable experience in teaching, examining, training and authoring in the field of Critical Thinking. They were tasked with the following: (1) to derive a Critical Thinking definition; (2) to derive a Critical Thinking taxonomy; (3) as far as possible, to map Cambridge Assessment qualifications against the taxonomy; (4) to identify skills closely related to Critical Thinking but which are not considered to be Critical Thinking.

A model or approach as to how to proceed through these tasks was required and the panel was offered three possibilities described below.

The top-down approach, working sequentially to derive first a definition as a group, then a taxonomy, followed by the mapping exercise, might be considered the 'pure' approach, in that the definition is derived before, and independent from, a consideration of the qualifications. However, an entirely pure approach in this respect may not be achievable: for the experts, their working knowledge of various CT assessments is implicit and likely to inform any work on the definition.

The bottom-up approach involves considering the Cambridge Assessment qualifications and assessments in some detail before deriving a definition. In one sense, this would be putting a framework around what was already existing, with the qualifications and assessments themselves providing the driving force for the activity. In other words, the bottom-up process might result in an overly self-confirmatory definition and taxonomy. However, this approach would have an advantage of 'reminding' the panel of (valid) aspects of Critical Thinking.

The iterative approach suggested was based upon the top-down model, where activities logically proceed from the definition. However, this model builds in a capacity to revisit and ultimately refine one step in the light of decisions about another step.

Unanimously, the experts chose to adopt the iterative approach. This proved a fruitful approach as, on occasion, the mapping exercise challenged the current version of the taxonomy: for example, the panel questioned whether one sub-skill should be presented as two separate sub-skills, or, conversely, whether two sub-skills were, in reality, inseparable and should be conflated.

The work on the definition and taxonomy provided Cambridge Assessment with a document and shared understanding of Critical Thinking. Certainly, the taxonomy provided a useful mechanism by which to map and quality assure extant assessments; to check their curricular relevance and domain coverage and hence provide one strand of evidence to support the inference that individuals who do well in Critical Thinking assessments demonstrate well developed skills in the Critical Thinking domain (Table 1).

Additionally, the taxonomy provided a way of ensuring the coherence and 'fit' of qualifications which were subsequently developed. The OCR Level 2 Thinking and Reasoning Skills qualification is an example of this, as the development process used the taxonomy throughout.

Critical Thinking is the analytical thinking which underlies all rational discourse and enquiry. It is characterised by a meticulous and rigorous approach.
As an academic discipline, it is unique in that it explicitly focuses on the processes involved in being rational.
These processes include:

- analysing arguments,
- judging the relevance and significance of information,
- evaluating claims, inferences, arguments and explanations,
- constructing clear and coherent arguments,
- forming well-reasoned judgements and decisions.

Being rational also requires an open-minded yet critical approach to one's own thinking as well as that of others.

Table 1
The Cambridge Assessment taxonomy of critical thinking.

	Skill/process	Sub skills/processes
1	Analysis	A Recognising and using the basic terminology of reasoning B Recognising arguments and explanations C Recognising different types of reasoning D Dissecting an argument E Categorising the component parts of an argument and identifying its structure F Identifying unstated assumptions G Clarifying meaning
2	Evaluation	A Judging relevance B Judging sufficiency C Judging significance D Assessing credibility E Assessing plausibility F Assessing analogies G Detecting errors in reasoning H Assessing the soundness of reasoning within an argument I Considering the impact of further evidence upon an argument
3	Inference	A Considering the implications of claims, points of view, principles, hypotheses and suppositions B Drawing appropriate conclusions
4	Synthesis/construction	A Selecting material relevant to an argument B Constructing a coherent and relevant argument or counter-argument C Taking arguments further D Forming well-reasoned judgements E Responding to dilemmas F Making and justifying rational decisions
5	Self-reflection and self-correction	A Questioning one's own pre-conceptions B Careful and persistent evaluation of one's own reasoning

This activity provided a key aspect of the validation of tests – a way of inspecting the extent to which the assessment is relevant to the content domain and thus providing an argument for supporting the adequacy and appropriacy of test score interpretations.

2.2. Going further with definitions – a glossary of Critical Thinking terminology

Having defined the overall content domain of Critical Thinking, it was clear we needed to go beyond this macro overview to a more micro, or fine-grained level, and examine the terminology and concepts within the content domain. Problems of definition seem to be more of an issue in Critical Thinking than in many discipline areas. There are a number of reasons for these definitional issues. There are 'false friends' whereby the meaning used in everyday usage might differ subtly – or more substantially – from a more technical usage (consider 'dilemma', 'reliability', 'argument'). Other concepts are closely related or overlapping (consider 'flawed' arguments versus 'unsound' arguments; 'circular' arguments versus 'begging the question'). And some things are easily mistaken for one another when trying to analyse thinking presented in normal everyday language–consider an 'argument' versus an 'explanation', which can share the same indicator words, such as 'because', 'since', 'therefore', etc.

For an assessment agency, these problems of definition create an interesting challenge. When writing and producing assessment material (such as specifications/syllabuses, question papers and mark schemes), clear, unambiguous communication is essential. Writing questions (or 'items') for tests that that use a closed response format (rather than requiring an open, extended response) require much precision, and the task can become very demanding if item writers do not have a

shared understanding of a concept. Another undesirable scenario emerges if an item and source material leaves too much scope for multiple interpretations by test takers. Once again, the important question of validity arises – in terms of item and test construction; validity, arguably, in its most basic form is whether a test is measuring what it claims to (Ruch, 1924).

The quality assurance processes around question paper production aim to remove as much ambiguity as possible. In order to help this process, we wished to create an authoritative text – a glossary – on all common terms encountered in Critical Thinking, to clarify terms as far as possible and where necessary, draw out differences between closely related terms. Again, Cambridge Assessment assembled a number of experts in the field to discuss, write and peer-review the entries. The first step was to determine which terms and concepts should be included in such a text, and which should be excluded – again, an exercise in mapping the territory of Critical Thinking. We decided to include some terms which, in the first definition and taxonomy exercise, we had decided were on the fringes or outside of the concept of Critical Thinking. Such terms included 'syllogism', 'problem solving' and 'statistical reasoning'. Their inclusion in the glossary is not because these are considered core CT skills, but rather because we believe understanding them can help to draw fine distinctions between closely related concepts and thus can benefit students of Critical Thinking, and anyone else wanting to improve their thinking skills.

As the glossary project progressed, it became apparent that we had a text that could have some value to a wider audience and not just people directly involved in creating assessment material. Teachers and students might be also very interested as well as academics and anyone wishing to sharpen their thinking. As a consequence, it was published under the title 'An A to Z of Critical Thinking' (Black, 2011) to reach a wider audience.

3. Critical Thinking AS/A level and schools

A levels (Advanced Levels) are the main academic track qualifications for second level education in the UK. Designed for 16–19-year-olds, to be completed after a two year course of study, they are offered in a wider range of subjects and are the standard academic route into further education. The A level comprises of two parts – 'AS' and 'A2'. AS (Advanced Subsidiary) level is normally completed after the first year of study and A2 after the second year of study. Together, they form an A level overall. Usually, candidates take three or four AS level courses in their first year of study, and continue with three in their second year of study in order to achieve overall three full A levels and one AS level. Each AS/A level is designed to be delivered as a separate course.

Critical Thinking was first introduced as an AS level in 2001 and as a full A level in 2006. At the point at which this second research study was conducted, candidate numbers nationally for CT entries for AS level had risen dramatically from just over 2000 candidates in 2001 to 22,000 candidates in 2008. Because if was such a new subject, few teachers were in a position to benefit from years of teaching experience. Furthermore, few teachers had any direct instruction in Critical Thinking in their own educational experience. We were interested to see how schools were managing to create a sufficiently expert workforce to deliver this course.

One reason why practices in schools were of interest was because of the relatively low proportion of candidates receiving the top grade in CT in comparison with other AS level subjects. There were three possible explanations for this:

(1) performance standards exhibited by candidates reflect a low level of teaching provision;
(2) performance standards exhibited by candidates reflect a low level of candidate motivation to achieve in this discipline;
(3) the discipline is highly demanding.

Such explanations, of course, are not mutually exclusive. The results could be due to a combination of all three. This particular study did not allow us to investigate explanation (3), which would require a comparability study of a different nature. Furthermore, it was not possible in the research study to entirely disentangle explanation (2) and explanation (3). However, it was possible to dig a little deeper and find out how schools were delivering the subject, how committed they were to resourcing the subject as well as a sense of the perceptions of Critical Thinking in schools.

This project took the form of a survey. All centres with entries for OCR CT AS/A level were contacted (ca. 1000) to ask if they would take part in a survey. Responses were received from 236 teachers representing just over 20% of all CT centres and 34.3% of the AS candidate entries.

A full explanation of results is available in Black (2009). In this paper, a brief overview is offered.

3.1. Results

Many teachers were strong advocates of teaching Critical Thinking. They referred to a range of benefits for students. A number of teachers highlighted the enjoyment they derived from challenging, stretching and encouraging their students' thinking:

> *It gives me the opportunity to challenge the brightest students and to develop their intellect far more than is possible at KS4.*

> *I like encouraging thinking – education should, as Hemingway put it, "make you a good crap detector"*

The majority of respondents reported that they positively valued (and that they also believed their students positively valued) Critical Thinking (see Fig. 2).

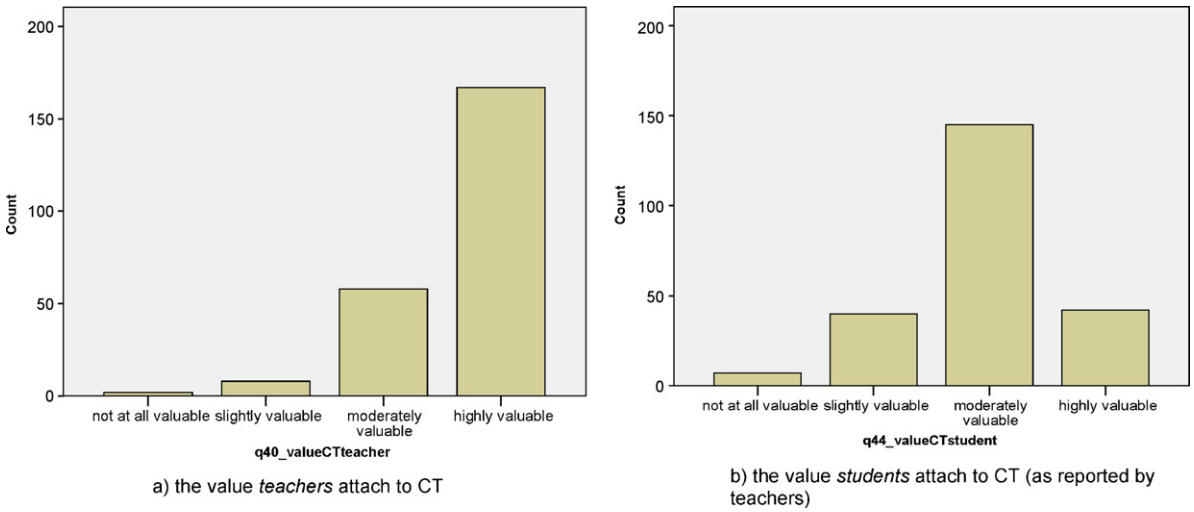

Fig. 2. The value attached to CT by both teachers and students (as reported by teachers).

As Fig. 2 shows, the modal response was that teachers find Critical Thinking *highly* valuable and this was supported by many interesting additional comments – some relating to everyday thinking; others about thinking in other academic subjects:

> *The pupils have poor analytical skills and believe most conspiracy theories and media headlines shown to them. They are reluctant to analyse what they read on the web in particular. This should be a growing concern and CT combats this to some extent.*

> *Its value lies primarily in that it is applicable to every other academic school subject I have come across. . . many that pupils will not encounter until university or later life.*

The vast majority of respondents (226/236) believed that CT benefitted students in their other subjects due to transferability of these skills.

> *Many subjects call for reasoned arguments. What better way to prepare them?*

> *. . .the majority [of students] find it quite useful and they now write better essays or think more logically. One said "It has changed my whole way of thinking".*

Additionally, teachers often believed they were benefitting, both personally and professionally, from teaching CT. This is an important collateral benefit.

> *Teaching this subject has altered the way I think. I find myself using the skills not only in the classroom but also in meetings and other aspects of my life.*

And many found delivering CT pushed and challenged their normal modes of teaching, adopting different teaching styles/pedagogical approaches. This finding echoes that of Baumfield and Oberski (1998) that (broader) thinking skills approaches in classrooms were popular with teachers because they foster different patterns of interaction in the classroom – more discussion, more listening to students; and as a result of this, more unpredictability and uncertainty in lessons.

> *It's a subject you can 'discuss'; it requires little didactic teaching which is good.*

> *I really like challenging myself in the teaching and sometimes I do not know the answers and work them out with the students. I find that very powerful as a teaching tool and model for learning.*

> *More unanswerable questions are raised than in other subjects and there is a real opportunity to challenge and explore each other's points of view.*

> *Every lesson is different – I am always surprised or stimulated by student responses.*

Some respondents clearly indicated that a more uncertain classroom placed them outside their comfort zone. Again, this echoes Baumfield and Oberski (1998) as well as Sternberg (1987).

> *It's not like teaching other subjects where you can hide being wrong or not knowing; students lose faith in your ability to teach; this has implications for the senior role I play in college.*

On the other hand, the 'different' approach to learning was perceived as something the students also enjoyed:

Students seem to value the subject as being different; they enjoy the immediacy of its challenged; they enjoy the way it enhances their ability to win arguments. . . they think it's "cool"

However, some of the enthusiasm for CT was tempered by other issues. Across the survey, there were a number of frustrations or disappointments reported. These are grouped (broadly) around the 'explanations' mentioned earlier.

There was some evidence supporting explanation (1). For the majority of the teachers (212/236 respondents), Critical Thinking was their second or third subject. Two thirds of respondents reported that Critical Thinking constituted less than 20% of their teaching timetable and some of these respondents made it clear it was a lower priority in terms of preparation, resourcing and training, for example:

. . . I do have the opportunity to attend INSET[7] *but my other subjects and classes take priority*.

The course, as an AS level, should require approximately 140–160 guided learning hours (i.e., contact teaching time). However, the mean teaching time indicated by respondents was just over 57 h. Teachers' ratings of the adequacy of this timetabling indicated that 60% of respondents believed they had about the right amount of time while 40% believed this was not enough time to deliver the course. Indeed, a common theme in the analysis of open questions was frustration with the senior management teams in schools who, the teachers believed, were not resourcing the course properly enough.

I would LOVE to teach Critical Thinking properly but I am not given the time of the timetable, the teacher resources, or the support I require in school either to teach my own classes properly or to co-ordinate the delivery of it school-wide.

When asked about training and resourcing, while the majority had attended one or more INSET sessions, many felt this, though useful, was insufficient training to become competent themselves in the discipline and deliver the course. Some teachers had apparently become very resourceful themselves, producing their own teaching materials and student resources.

I've had no support – I've done it all myself.

I feel there is a desperate need for far more training for people who, like me, are 'flung in at the deep end' and have little clue of what they are expected to deliver!

There was some evidence supporting explanation (2) – lack of student motivation. This was an interesting area because a very mixed picture emerged. Broadly speaking, the centres fell into two categories: those for which Critical Thinking was an optional (extra) AS/A level; and those who made Critical Thinking mandatory for all students. In the first category, where students can make a positive choice for something, motivation is likely to be good – and this was strongly evidenced in the survey findings. In the second category, where it was imposed upon the students (and the staff), motivation is likely to be low. There was strong evidence in the survey for this pattern. Low motivation also interacted with limited timetabling.

It only takes a look at the timetable for students to make up their mind about how valuable the subject is in comparison with other subjects.

Whilst the majority of the students see it as beneficial, there are some that resent having to do a compulsory subject once they are in the Sixth Form.

Finally, one quotation illustrates an interaction between quality of the teaching and student motivation:

As they [students] don't choose this subject as a main AS they give the subject very low priority. Interest in the subject depends very much on who is teaching it. Often staff who are uninterested in the subject are asked to teach CT to fill their timetable.

As for explanation (3), it was not a key purpose of the study to explore this. However, some respondents certainly expressed a *perception* that Critical Thinking is a difficult subject – both in terms of the challenge of mastery of the skills and in terms of attaining a good grade.

Students enjoyed the course. While many found it intellectually challenging and may come out with low grades there was a real sense of achievement for completing the year. All students felt it helped them in other subject areas.

3.2. Discussion of CT in schools survey

A fuller discussion can be found in Black (2009). The survey revealed an interesting mixture of positive and negative practices and perceptions surrounding CT. A picture emerged of teachers who were committed to CT and up-skilling themselves despite lack of resourcing from their senior management, with motivated students who made a positive choice to take the course and were keen for a challenge. But there was also evidence of lack of teaching time, disaffected and frustrated teachers with students who were obliged to take the course. Certainly, there was some evidence to lend support to hypotheses (1) and

[7] INSET stands for 'In Service Education and Training'. INSET often takes the form of a one-day course.

(2). However, one positive collateral benefit reported was of teachers who developed professionally as a result of teaching Critical Thinking to the benefit of teaching in their other (main) subjects.

4. The impact of Critical Thinking AS/A level on wider academic attainment

As reported above, the vast majority of teacher respondents to the survey believed that students who studied Critical Thinking benefited in terms of their other AS/A level performance. We wanted to investigate this further and see whether there was any statistical evidence to support these perceptions.

It is worth remembering that, in the UK, the teaching of Critical Thinking AS level takes place as a separate course. In terms of the wider context of teaching Critical Thinking, there has been much discussion and debate given to the most effective model of teaching CT – as a separate discipline or infused into other subjects. Abrami et al. (2008), in an excellent meta-analysis, indicates that the most effective teaching of CT takes place when it is explicit *and* in the context of other subjects (the 'infusion' approach) rather than in a free-standing course of its own. This raises the question of whether, in the UK context. There is still value in delivering CT AS level as a discrete subject, given that we might think that the greatest value of CT lies precisely in its ability to increase rational thought in a whole range of academic and non-academic domains.

For this study, we conducted two analyses (for further details, see Black & Gill, 2011). In the first analysis, two groups of students of more than 2000 UK students were compared. One group comprised all those candidates ($n = 2208$) who had a good grade in CT AS level (grade A or grade B); while the other group (also $n = 2208$) had not taken CT. The groups were matched for prior attainment at GCSE. A comparison of the groups' overall mean score showed that the CT groups' A level performance was higher than the non-CT group – equivalent to about one grade difference across three A level subjects. This was highly significant (Kolmogorov–Smirnoff test; $p < 0.001$).

Using the same sample of two groups, we also investigated performance at A level in the most popular subjects individually. Briefly, this provides evidence that candidates who achieved high grades in AS level CT performed better overall in A level subjects (across the science and humanities divide) than those who did not study CT at all. There is evidence that this advantage presents itself across a wide range of subjects – in sciences, social sciences, humanities and arts subjects. Table 2 provides an overview of the results for this analysis.

In the second strand, we conducted a regression analysis – this time comparing candidates with AS Critical Thinking (all grades) with those without Critical Thinking whilst also controlling for the prior attainment (GCSE mean score) of candidates. This is a more complex procedure and full description is available in Black and Gill (2011). In summary, this analysis gave further evidence that candidates taking the AS level in CT and getting a grade D or above, overall showed increased A level outcomes.

Some final thoughts on this study. As it is not a randomised control study (i.e., with students randomly allocated to different groups), what this study cannot take into account is whether there are some other differences between the two groups of candidates (CT versus non CT) which may explain the difference in A level performance. For example, there may be some sort of school effect that we could not identify – for example, perhaps only 'better' schools offer CT.

Furthermore, although we have shown an association between taking CT at AS level and increased performance at A level, we cannot, from this analysis, be sure the former *causes* the latter. It may be that candidates who perform well on CT do so because they already (independently) possess the skills and attributes to perform well in other academic subjects. However, this explanation is not entirely plausible to account for the results of the study, as these notional skills and attributes had not differentially benefited candidates at GCSE level.

If we accept the interpretation that studying CT AS level can improve performance in other subjects, it is worth reflecting upon this in conjunction with the survey findings outlined in Section 2 of this article. According to this survey, many candidates taking CT were not likely to have an optimal teaching and learning experience due to lack of resourcing and training around its delivery. Thus, to see any significant effect at all in this context could be interpreted as an unexpected and positive finding. It is a reminder that there is some broader purpose to teaching CT in schools – adding value to students and enhancing their performance in other academic subjects. Further research would be of interest to discover under what conditions CT skills are most easily transferred to other subjects when it is delivered in a standalone model.

Table 2
Comparing mean A level grade score for CT group and non CT group.

	Mean A level grade	
	Non CT candidates	CT candidates (grade A or B)
Overall – all subjects	8.68	9.12
Biology	8.76	9.17
Chemistry	8.96	9.35
Physics	8.94	9.33
Maths	9.02	9.32
Geography	8.64	9.17
Economics	8.98	9.46
Psychology	8.01	8.55
English	8.91	9.24

Table 3
Correlation coefficients between TSA scores and 1st year examination performance.

		TSA problem solving score	TSA Critical Thinking score	TSA total score
Degree course	Computer science	0.227**	0.216**	0.268**
	Economics	0.275**	0.205**	0.281**
	Engineering	0.278**	0.182**	0.268**
	Natural sciences	0.193**	0.128**	0.189**

Source: Emery (2007).
** Significant at the 0.01 level.

Table 4
Guidelines for interpreting correlation coefficients in predictive validity studies.

Validity coefficient	Interpretation
Above 0.35	Very beneficial
0.21–0.35	Likely to be useful
0.11–0.20	Depends on circumstances
Below 0.11	Unlikely to be useful

Source: US Department of Labor, Employment Training and Administration (1999).

5. Predictive validity of admissions tests

The purpose of admissions tests is to provide a higher education institution with some information about the candidate which is helpful in predicting the success of that candidate on their chosen course. Therefore, there is a burden upon the test provider to investigate the relationship between performance on the admission test and subsequent university performance.

> "Validation studies should be commissioned, where necessary . . . For example, if the assessment is intended to predict later performance in a following stage of education or training, the relationship between the outcomes of the assessment and the outcomes of the subsequent education and training programme should be systematically explored." Cambridge Assessment (2009, p. 9)

Examining the predictive power should be considered a prime validation activity of such tests. Cambridge Assessment has two key admissions tests which contain CT items – the TSA and BMAT. In both cases, CT is an element of the test. Higher Education institutions who use these tests may have slightly different processes around how they use information from these tests. Some may use it as a 'hurdle' (i.e., a candidate has to have a score >x in order to be given further consideration for the course), while some may use it as an additional piece of evidence in their selection processes alongside other pieces of evidence provided in the application process such as other qualifications and grades, a personal statement and references. In the former case, a 'good' score on the test is a necessary condition for admission (though not sufficient); and in the latter, it is neither necessary nor sufficient. But, regardless of which strategy a HE institution chooses to adopt, there needs to be some sense of the predictive validity of the tests if they are to have any value in the admissions process.

Cambridge Assessment routinely runs such studies (e.g., Emery, 2009; Emery, Bell, & Vidal Rodeiro, 2010). Such analyses examine the relationship between the admissions test score and subsequent degree performance, e.g., in the first year of university study or in the final year. These analyses are usually one of two types: (1) correlational – between the admissions test and marks and (2) logistic regression – using admissions test scores and overall degree class. Table 3 and Figs. 3 and 4 (all from Emery, 2007) show some example results from such analyses.

Interpreting correlational analyses for predictive validity is not always straightforward. This is because the candidates who have been accepted for the course tend to have a higher mean and narrower range of test performance than TSA candidates overall. This problem is referred to as 'range restriction' as it restricts the magnitude of the correlation coefficients. While some statisticians may use corrective formulae for range restriction, in general the authors prefer to use the 'uncorrected' coefficients and interpret them using guidance (see Table 4). For a fuller discussion of predictive validity analyses, see Bell (2007).

Therefore, looking at Tables 3 and 4 in conjunction, we can see that the coefficients fall mainly into the 'likely to be useful' category. Interestingly, there are some differences between the different degree courses and this possibly reflects the extent to which such thinking skills (as measured in the TSA) are subsequently rewarded when compared with other skills, attributes and knowledge required of the course.

Logistic regression predicts the probability of a binary outcome such as an event occurring or not (e.g., obtain a first versus not obtain a first) as a function of a predictor variable (e.g., score on an admissions test).

In a logistic regression plot, the strength of the predictive relationship can be discerned from looking at the steepness of the slope. A flat line indicates the predictor has no relationship to the outcome variable. A steep line indicates the predictor variable has a strong ability to predict the outcome variable.

These examples of logistic regression plots show that TSA is a strong predictor of performance on Engineering and Economics (Figs. 3 and 4). We can see (Fig. 3) that the higher the score on both parts of the test, the higher probability of achieving a first class outcome in the first year. There total score for both parts of the TSA gives the strongest prediction.

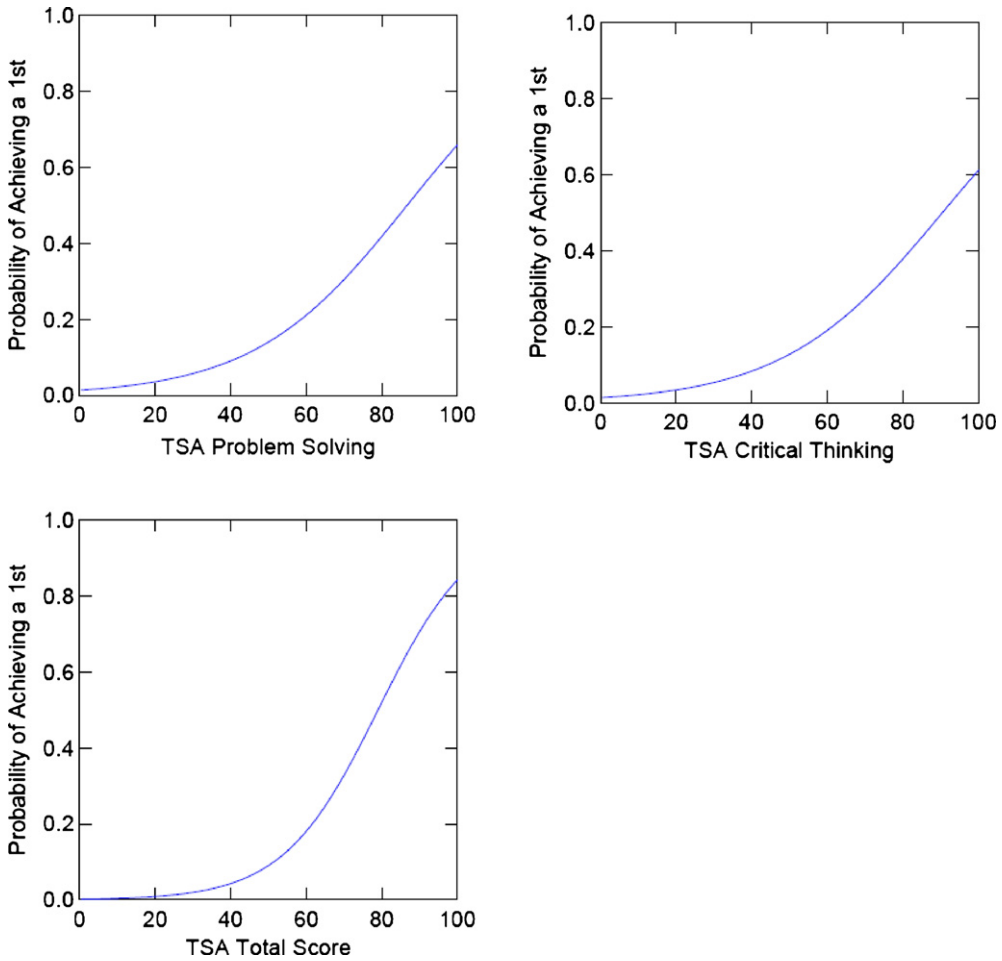

Fig. 3. Logistic regression plots showing the probability of achieving a 1st class outcome in Economics (first year) as a function of TSA scores (observations = 145, 1st = 36).
Source: Emery (2007).

Fig. 4 shows again a strong predictive relationship whereby low scores on the TSA are associated with high probability of achieving a 3rd class outcome or below (failure or withdrawing) in the first year.

Similar analyses have been produced for TSA and other degree courses, as well as for the BMAT showing similar outcomes.

Such analyses mean that we can have reasonable confidence in the inferences which test users make about scores on this test.

6. Discussion

The programme of research described is quite diverse in purpose, method and scope. Discussions of individual research activities can be found within the individual papers as referenced. The discussion here attempts to provide a commentary on the work as a whole.

What ties all of these projects together is validity in the sense of educational measurement – that we should have tests and qualifications which are valid by design (Misley, 2007), and that we should acquire "empirical evidence and theoretical rationales [to] support the adequacy and appropriateness of inferences and actions based on test scores" (Messick, 1989). The definition, taxonomy and glossary work help support this from the point of view, ensuring that qualifications and tests are sampling from the Critical Thinking domain and therefore that inferences made from test scores about candidates' ability in Critical Thinking are sound.

The survey work around the delivery of Critical Thinking in schools also has a validity aspect to the work. The survey of CT teachers helps us understand the behaviours associated with this specific assessment. When an assessment agency designs a qualification, one design principle is that there should be a clear need for the assessment to benefit the individual, society and/or the economy (The Cambridge Approach, 2009). In other words, the test/qualification should promote something of value. Frequently, in the world of assessment, the design may be predicated on this principle, but it may be something which

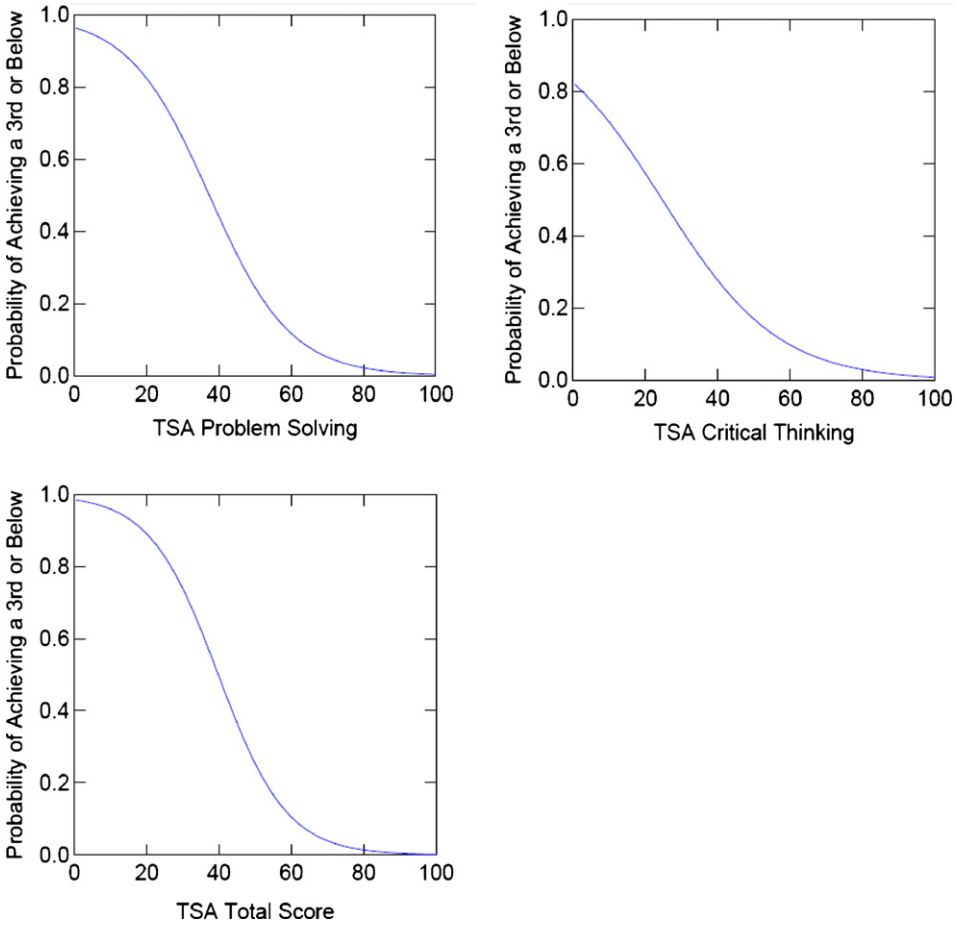

Fig. 4. Logistic regression plots showing the probability of achieving a 3rd class outcome or below in Engineering (first year) as a function of TSA scores (observations = 370, 3rd or below = 37).
Source: Emery (2007).

is not re-visited to see whether this principle is borne out in practice. The results showed that, despite patchy resourcing of the CT courses in schools, that teachers and students (on basis of teacher reports) very much valued this subject. They found value in the skills the qualification promotes, not only for other academic subjects, but also for navigating through uncertainty in everyday life. In addition, there is strong statistical evidence that studying CT also benefits students' achievements in other subjects. This is not to say that the work of an assessment agency ends here and that we can be happy the qualification does its job. There will continue to be evaluations and refinements to help ensure that future AS/A level Critical Thinking specifications continue to be relevant and add value to the students' learning experiences and school achievements.

For those tests, such as BMAT and TSA, whose purpose is for selection and prediction of future performance, investigations into the relationship and subsequent performance are an essential aspect of validation. Overall these studies indicate that scores on these tests are likely to positively add to information about a candidate so that admissions tutors can make more informed choices.

Finally, it is worth standing back from the 'impact' research for a moment and remembering that the impetus to launch qualifications in CT came from the Informal Logic movement (e.g., Fisher & Scriven, 1997). Their purpose was to improve ways of thinking when using natural language, i.e., outside of the environment of formal logic, and to promote critical thinking within other domains. The finding that students studying CT AS level tend to perform better on their other A levels provides some evidence to support the idea that CT, even when taught as a standalone subject, can promote skills and enhance academic achievements across a wide variety of other domains.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., et al. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, *78*, 1102–1134.

American Educational Research Association, American Psychology Association & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Baumfield, V. M., & Oberski, I. M. (1998). What do teachers think about thinking skills? *Quality Assurance in Education*, *6*, 44–51.
Bell, J. F. (2007). Difficulties in evaluating the predictive validity of selection tests. *Research Matters: A Cambridge Assessment Publication*, *3*, 5–10.
Black, B. (2008). Critical Thinking—A definition and taxonomy for Cambridge Assessment: Supporting validity arguments about Critical Thinking assessments administered by Cambridge Assessment. In *Paper presented at 34th international Association of Educational Assessment annual conference* Cambridge, 9th September 2008.
Black, B. (2009). Introducing a new subject and its assessment in schools: The challenges of introducing Critical Thinking AS/A level in the UK. In *Paper presented at the Association of Educational Assessment-Europe conference* Attard, Malta.
Black, B. (Ed.). (2011). *An A to Z of critical thinking*. London, New York: Continuum Books.
Black, B., & Gill, T. (2011). Does doing Critical Thinking AS level confer any advantage for candidates in their performance on other A levels? *Research Matters. A Cambridge Assessment Publication*, *11*, 20–25.
Cambridge Assessment. (2009). *The Cambridge approach. Principles for designing, administering and evaluating assessment*. Available online at: http://www.cambridgeassessment.org.uk/ca/digitalAssets/181348_cambridge_approach.pdf Accessed 12.01.12
Dewey, J. (1909). *How we think*. Boston, MA: D.C. Heath and Co.
Ebel, R. L. (1965). *Measuring educational achievement*. New Jersey: Prentice Hall.
Emery, J. L. (2007). *The predictive validity of the Thinking Skills Assessment: A combined analysis of three cohorts*. Available at: http://www.admissionstests.cambridgeassessment.org.uk/adt/digitalAssets/121139_combined_TSA_validity_report1.pdf
Emery, J. L. (2009). The predictive validity of the BioMedical Admissions Test for pre-clinical examination performance. *Medical Education*, *43*.
Emery, J. L., Bell, J. F., & Vidal Rodeiro, C. L. (2010). The BMAT for medical student selection—Issues of fairness and bias. *Medical Teacher*, *33*, 62–71.
Ennis, R. H. (1996). *Critical thinking*. New York: Prentice Hall.
Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Executive summary. The Delphi report*. Millbrae, CA: California Academic Press.
Fisher, A., & Scriven, M. (1997). *Critical thinking: Its definition and assessment*. Norwich: Centre for Research in Critical Thinking.
Glaser, E. (1941). *An experiment in the development of critical thinking*. New York: Teachers College, Columbia University.
Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd edition, pp. 13–100). Washington, DC: American Council on Education.
Misley, R. J. (2007). Validity by design. *Educational Researcher*, *36*, 463–469.
Paul, R. (1992). Critical thinking: What, why and how? *New Directions for Community Colleges*, *20*(Spring), 1.
Ruch, G. M. (1924). *The improvement of the written examination*. Chicago: Scott, Foreman and Company.
Sternberg, R. J. (1986). *Critical thinking: Its nature, measurement and improvement*. ERIC Document Reproduction Service ED272882.
Sternberg, R. J. (1987). Teaching critical thinking: eight ways to fail before you begin. *Phi Delta Kappa*, *68*, 246–249.
US Department of Labor, Employment Training and Administration. (1999). *Employer's guide to good practices*. Accessed at: http://www.chrysaliscorporation.com/pdf/Testing_and_Assessment_Guide.pdf

This document was initially designed for print and as such does not reach accessibility standard WCAG 2.1 in a number of ways including missing text alternatives and missing document structure.

If you need this document in a different format please email admissionstesting@cambridgeassessment.org.uk telling us your name, email address and requirements and we will respond within 15 working days.