

Measuring Critical Thinking

Abstract

This article discusses examples of attempts to measure critical thinking, from the beginning of critical thinking literature to the present day. Assessments are discussed from academic and professional settings. The limitations of these assessment-based measurements are considered, as well as the potential ways of improving them. The difficulty in quantifying an abstract concept like critical thinking is discussed. Finally, this article discusses the role of Endoxa Learning in developing and assessing critical thinking skills and how it tackles some of the limitations of other methods of measuring critical thinking.

Why Should we Measure Critical Thinking?

Critical Thinking is an Objective of Education

Critical thinking is a key part of education, whether it is explicitly mentioned within the curriculum or not. The Framework for Higher Education Qualifications (FHEQ), includes the following descriptors which are specific to critical thinking:

Level	Outcomes / Abilities involving critical thinking
4	<p>Communicate the results of their study/work accurately and reliably, and with structured and coherent arguments</p> <p>An ability to present, evaluate and interpret qualitative and quantitative data, in order to develop lines of argument and make sound judgements in accordance with basic theories and concepts of their subject(s) of study</p>
5	<p>Effectively communicate information, arguments and analysis in a variety of forms to specialist and non-specialist audiences and deploy key techniques of the discipline effectively</p> <p>Knowledge of the main methods of enquiry in the subject(s) relevant to the named award, and ability to evaluate critically the appropriateness of different approaches to solving</p> <p>Knowledge and critical understanding of the well-established principles of their area(s) of study, and of the way in which those principles have developed</p> <p>Use a range of established techniques to initiate and undertake critical analysis of information, and to propose solutions to problems arising from that analysis</p>
6	<p>Conceptual understanding that enables the student to devise and sustain arguments, and/or to solve problems, using ideas and techniques, some of which are at the forefront of a discipline...</p> <p>Critically evaluate arguments, assumptions, abstract concepts and data (that may be incomplete), to make judgements, and to frame appropriate questions to achieve a solution - or identify a range of solutions - to a problem</p>

red = argument-related, blue = critical values

Table 1: FHEQ Qualification Descriptors for the UK, excluding Scotland

This illustrates the importance of critical thinking for academic success in higher education.

In the USA, critical thinking is actively embedded within the curriculum at a secondary school level. The US Common Core learning objectives for English Language Arts Standards, Grade 9-10, shows this emphasis on evidence of critical thinking skills:

- “Write arguments to support claims in an analysis of substantive topics or texts, using valid reasoning and relevant and sufficient evidence.”
- “Introduce precise claim(s), distinguish the claim(s) from alternate or opposing claims, and create an organization that establishes clear relationships among claim(s), counterclaims, reasons, and evidence.”
- “Provide a concluding statement or section that follows from and supports the argument presented.”.

These objectives illustrate the requirements for students to learn to make strong arguments and to be able to structure them in a way that leads smoothly to support a conclusion. Including critical thinking in the secondary curriculum is important, since it follows the recommendation of the Delphi Report (Facione, 1990) that critical thinking should be taught prior to higher education in order to be effective. This makes it important to include within learning objectives and assessment criteria, as a measure of academic achievement.

Employers Look for Critical Thinking Skills

Identified by the World Economic Forum as one of the “top 10 skills of 2025”, critical thinking is also essential for professional success (World Economic Forum, 2020). As a result, it is no surprise that a number of prominent employers test critical thinking skills as part of their recruitment process. The Watson Glaser Test, discussed later in this article, has been used by the Bank of England and the Government Legal Service among other employers (Feldman, 2021). Given that critical thinking has an increasing role in employability, it makes sense that critical thinking should be taught and tested among students in education.

However, critical thinking is an abstract and complex term, which makes it more difficult to measure and assess than other essential skills like reading and writing. The recognition that critical thinking is an important skill has led to the development of a number of assessment methods, although all of these methods have considerable limitations. Some of these methods, their strengths and their limitations are discussed in the following paragraphs.

Methods of Measuring Critical Thinking

1. Traditional Measurement of Critical Thinking in Education

Traditionally, a student's ability to think critically has been measured through essay-based assessment within humanities, arts and social sciences subjects. As a result, critical thinking ability is only an aspect of the assessment: students are simultaneously marked on their subject knowledge, reading and writing skills. Therefore, providing sufficient feedback about each aspect of the assessment is incredibly time intensive for teachers. With limited feedback from teachers, students may experience a cycle of writing essays and receiving negative feedback, without fully understanding why their argument is weak or how it can be improved.

"My lecturer used to write 'be more critical' on my essays, but I didn't know what he meant."

This can lead to confusion and result in limited improvement in later essays as a result. For students who struggle with reading and writing, this challenge is only exacerbated. Furthermore, this method fails to produce a separate, useful measurement of critical thinking that teachers can use to identify problems and drive improvement.

2. Essay Tests

The Ennis-Weir Critical Thinking Essay Test was one of the earliest attempts at measuring critical thinking separately from school subjects (Ennis and Weir, 1985). This test is relatively self-explanatory. Students are expected to read a passage of text and then write an essay evaluating the strength of the argument in the text. Structurally, the Ennis-Weir test is very methodical. A paragraph of evaluation is written for each paragraph of the source text, identifying strengths and weaknesses. The manual for this test provides detailed information on each paragraph, to aid an examiner in marking the student's essay. Ennis (1996) later wrote one of the most widely used definitions of critical thinking¹, so he is certainly an expert on the subject. This test measures students' ability to write critically and in depth about an unfamiliar subject.

However, the Ennis-Weir test has two main flaws, which arguably produce an unfair measurement of some students' critical thinking skills: the test requires good reading and essay writing skills; and the teacher is encouraged to mark answers subjectively (Ennis and Weir, 1985). As a result, the weaknesses of the Ennis-Weir test are much the same as those of the traditional method discussed previously. The authors emphasise that the examining teacher should have experience in critical thinking at an academic level and that they should practice the test prior to setting it for students. However, this only

exacerbates an existing problem; the essay test is as time consuming for teachers as a standard essay-based assessment. Furthermore, requiring strong reading and essay writing skills goes against the “Fairness” principal of the Delphi Report on critical thinking assessment. Based on this report, all critical thinking assessments “should not unfairly disadvantage or advantage groups of students on the grounds of reading ability” (Facione, 1990).

3. Critical Thinking as an A Level Subject

The momentum for assessing critical thinking skills reached a peak in the UK with the introduction of school subjects dedicated to critical thinking. A number of exam boards included a dedicated A Level subjects in the early 2000s, but only Cambridge International’s “Thinking Skills” curriculum remains. For this A Level, students are examined using a familiar exam format, with a mixture of short answer questions and essay style questions. Two papers, “critical thinking” and “applied reasoning” assess critical thinking, while the others focus on problem solving, with a greater focus on numerical questions.

Shorter questions in the critical thinking papers test aspects of a source material including:

- Identifying conclusions.
- Identifying when arguments contradict each other.
- Identifying strengths and weaknesses of arguments.

The longer answer question is similar to the Ennis-Weir test. It asks the student to write an essay using the source material. However, it also has differences. Firstly, a statement is introduced, which the student must support or challenge. Furthermore, rather than writing one paragraph per paragraph of source material, the student is expected to structure using clear “strands of reasoning”, comparing multiple source materials and make a “persuasive argument” (Cambridge International, 2019, p. 6).

Studying the course is argued to improve a number of skills among students, including:

- “A facility for independent thinking in the real world”
- “A facility for making well-reasoned judgements and decisions”
- “A transferable set of critical thinking, reasoning and problem-solving skills which are essential for success in higher education and employment”.

These abilities summarise effectively the benefits of critical thinking in education. Undoubtedly, spending a significant amount of time dedicated to a course on critical thinking means that students have more time to develop these skills.

The Cambridge International exams have clear advantages over the Ennis-Weir test, but they also share some of its weaknesses. Namely, although the verbal loading is reduced in the case of the shorter questions, strong reading and writing abilities are still necessary and this heavily impacts the criteria for achieving high marks. There is also a high workload for the examiner and for schools, who must choose to teach this subject over more traditional or accessible subjects. For the student, studying this A level means committing to a 2-year course of study. But UK universities do not accept it as part of their entrance requirements as they want to see students demonstrate their critical thinking within the traditional subjects. For these reasons, critical thinking as a subject has failed to make it into the curriculum in most schools.

"I would LOVE to teach CT properly but I am not given the time on 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."

- Teacher quote from Cambridge Assessment

4. Critical Thinking Tests for Employment

Some critical thinking tests are designed as a way of evaluating job applicants or of testing employee skills, rather than to test structured learning of critical thinking. As a result, these tests are short form and not based on a formal understanding of critical thinking theory.

The Watson Glaser test is perhaps the most well-known critical thinking test of this kind and is used most commonly by law employers. Questions test on "inference", "recognition of assumptions", "deduction", "interpretation" and "evaluation of arguments" (Pearson, 2002). The example below is from the "deduction" section:

Statement: No responsible leader can avoid making difficult decisions. Some responsible leaders dislike making difficult decisions. Therefore:	Proposed conclusions: 9. Some difficult decisions are distasteful to some people. 10. Irresponsible leaders avoid things they dislike. 11. Some responsible leaders do things they dislike doing.
---	--

Figure 1: Deduction question from a Watson Glaser practice paper (Pearson, 2002).

The short nature of the questions on the Watson Glaser test means that verbal loading is somewhat reduced, while the multiple-choice answers make marking easier than for essay-based assessments. However, the questions are necessarily simple, only testing on short statements. This means that the test does not measure proficiency in understanding and thinking critically about longer, more complex arguments. The questions are also based on scenarios that may not be familiar to exam takers.

Other short form employment tests include critical thinking questions similar to the Watson Glaser test, but these are often limited to a single section. The remaining sections are based on numerical or spatial reasoning, so the testing of critical thinking skills is even more limited. Tests with this format include:

- The General Reasoning International Test (GRIT)
- Critical Reasoning Test Battery (CRTB)
- Logiks General

These tests have similar structures and so their flaws are similar to those of the Watson Glaser test. However, the benefit of all of these multiple-choice assessments is that they can be transformed easily into statistics for the quantitative measurement of critical thinking among a sample group.

Summarising the limitations of critical thinking assessments

Although first published almost 40 years ago, this quote from Sternberg (1986, p. 27) encapsulates the ongoing challenges of measuring critical thinking through assessments:

“Many existing tests tend to be highly “verbally loaded,” and indeed, what these tests measure is not clearly distinguishable from verbal intelligence as it is traditionally operationalized in standard tests of intelligence. Moreover, there is a large gap between the ability to apply critical thinking in fairly trivial, highly structured, and usually multiple-choice tests, on the one hand, and in one’s everyday life, on the other. None of the tests came even close to bridging this gap.”

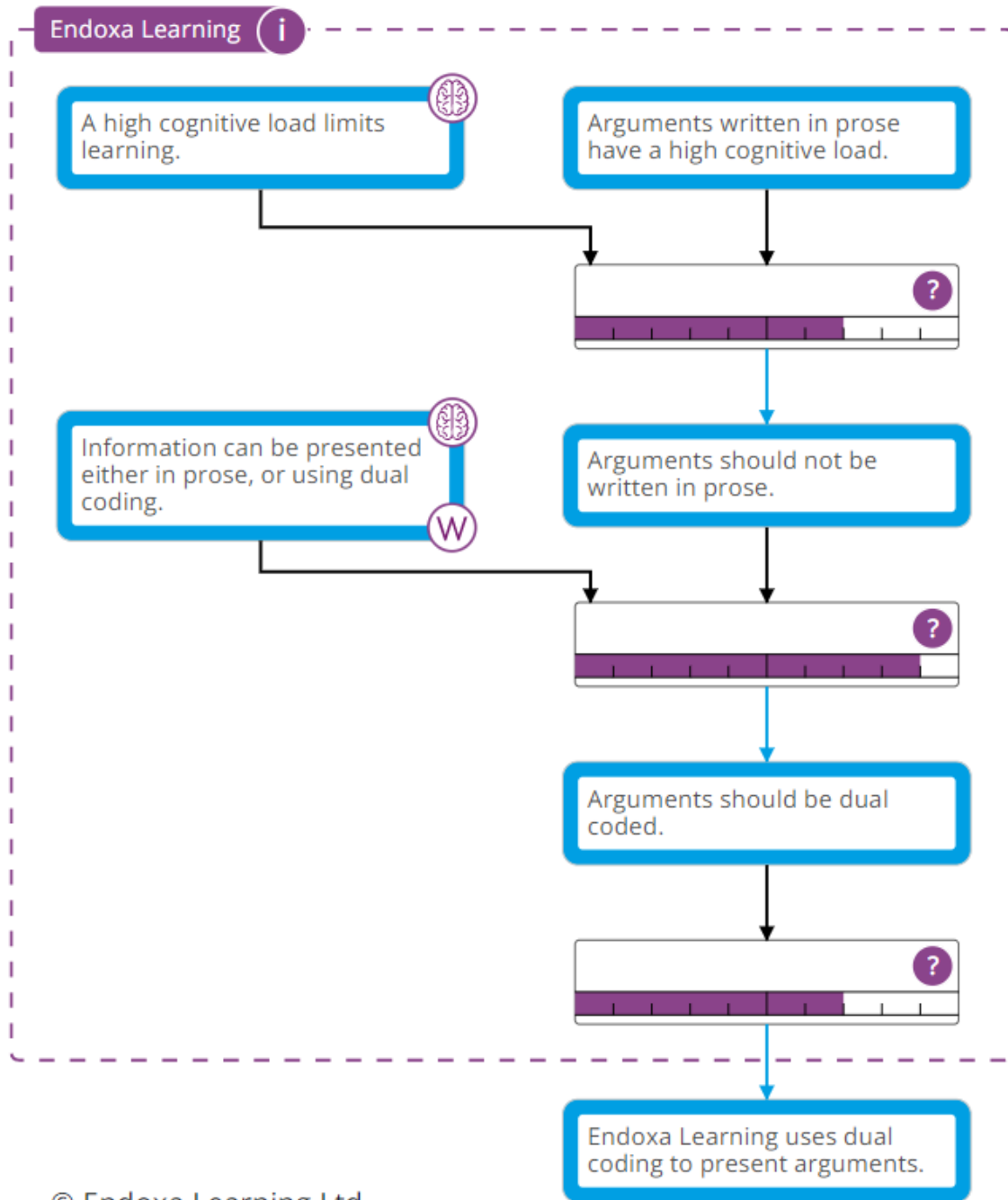
Improving the Measurement of Critical Thinking

In this section, we consider how the limitations of critical thinking assessments can be improved.

1. Reducing Verbal Loading

Critical thinking assessments should measure critical thinking skills. As discussed above, essay-based assessments have the unfortunate side-effect of measuring reading and

writing skills, which contradicts the components of a good critical thinking test as defined by the Delphi Report (Facione, 1990). Importantly, this negatively impacts exam takers who struggle with reading or writing. To make critical thinking measurement more accurate and to avoid disadvantaging some test takers, assessments should use limited words and simplified vocabulary.



© Endoxa Learning Ltd.

Figure 2: An argument graph explaining why Endoxa Learning uses dual coding.

At Endoxa Learning, we achieve this by **dual coding** arguments, displaying them as argument graphs (Figure 2). By mapping arguments visually and replacing complex sentences with arrows and boxes, we make it easier to understand the structure of an

argument. Testing includes hiding statements from the graph and asking the student to choose the correct answer, as well as quizzing them on their understanding of the overall argument. Unlike in essay-based assessments, students can work out an argument's strengths and weaknesses using its visualised structure, without wading through complicated prose.

2. Testing on Arguments of a Range of Complexities

While short form tests reduce the necessity for strong reading and writing skills, they do so at the expense of argument complexity. To measure critical thinking effectively, tests must measure participants' understanding of longer arguments that contain more controversy or have more than one branch.

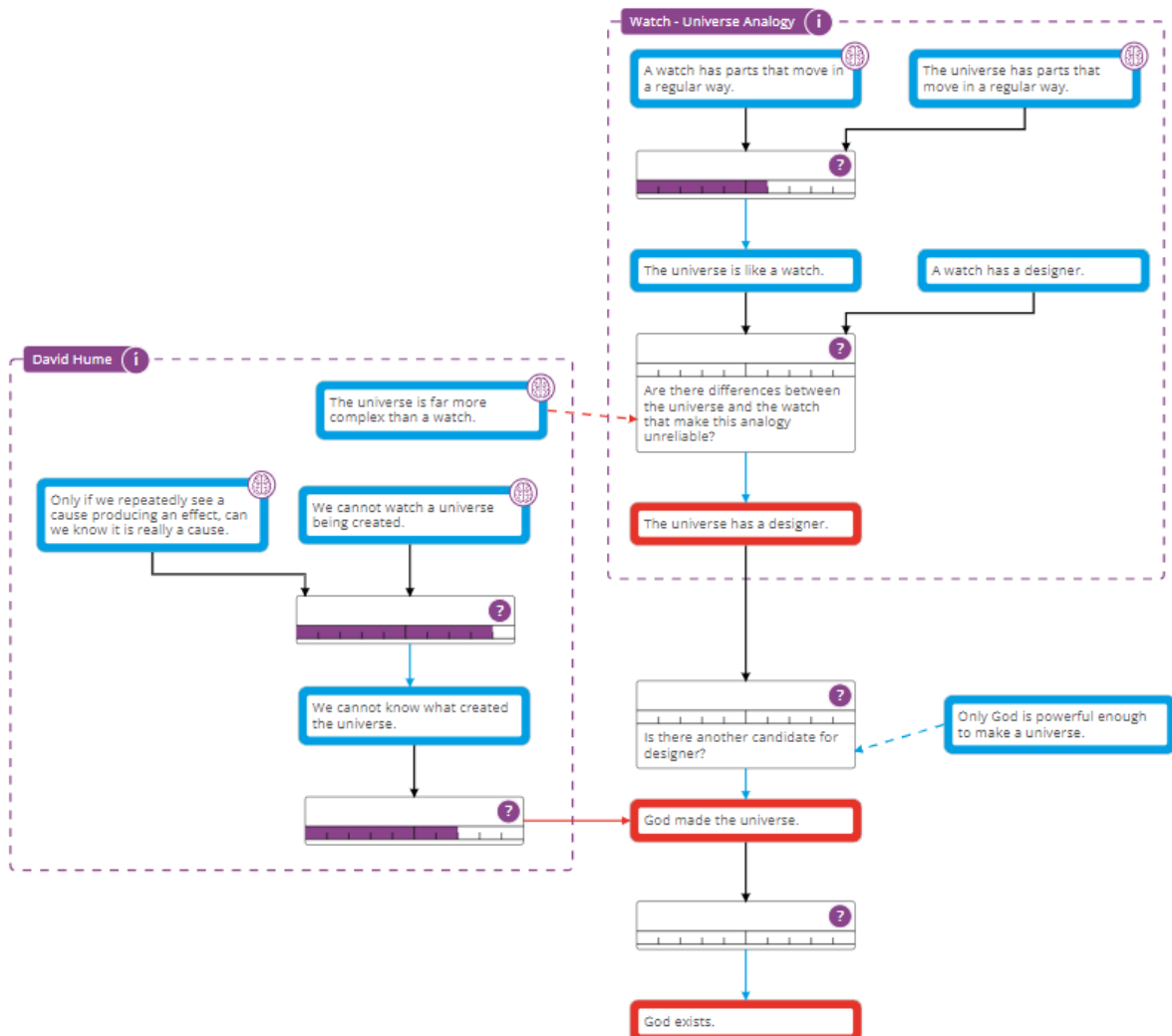


Figure 3: An argument graph showing a complex argument about whether or not God exists. Theoretical knowledge from David Hume is used to defeat the argument that God exists, allowing the user to turn the conclusion red to show that it is false.

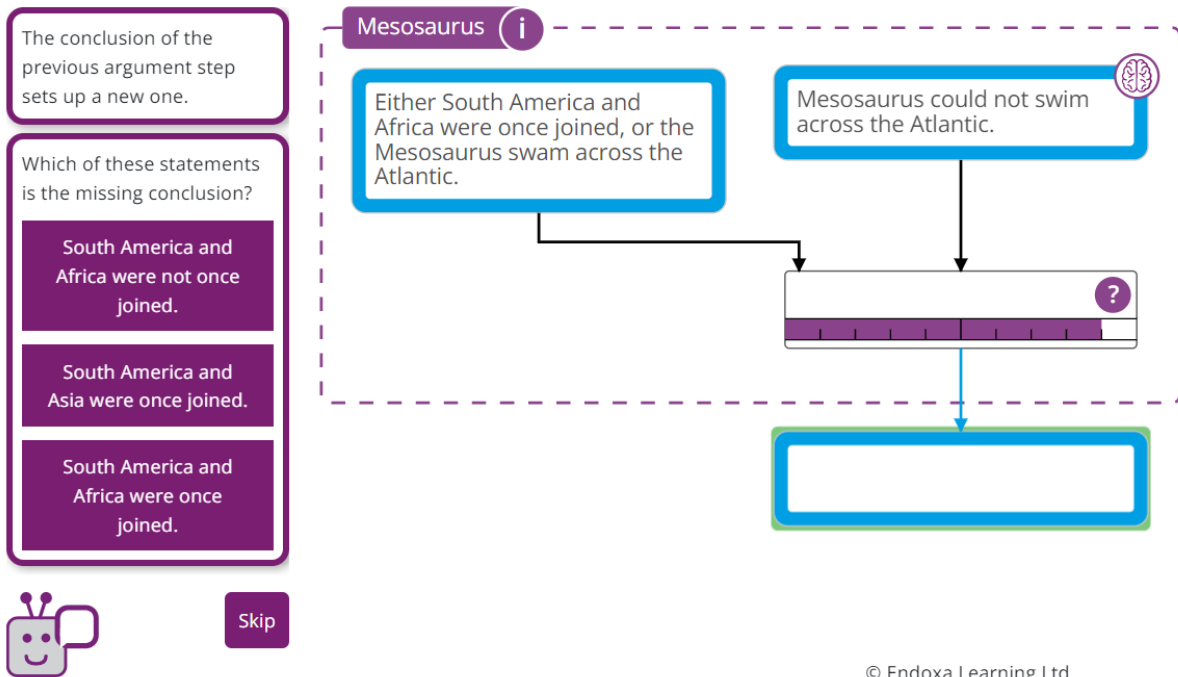


Figure 4: A single argument step from an argument graph. The task asks the student to identify the missing conclusion.

At Endoxa Learning, we test understanding of a range of arguments, with varying complexity. These can be a few short steps, or a much larger argument that includes different theories and case studies (Figure 3). This sets argument graphs apart from short form assessments, which cannot include complex arguments like the existence of God, climate change or other important debates. We automate testing of students' understanding of longer arguments, which means that checking an argument graph is much quicker than marking a traditional essay. As a result, the marking workload for teachers is reduced significantly. Additionally, Endoxa Learning can still test on shorter sections of an argument since each argument step is introduced individually (Figure 4).

3. Testing on Relevant Content

Whether essay-based or multiple-choice, the critical thinking tests discussed in the previous sections share the flaw of relying on irrelevant scenarios, with which test takers may be unfamiliar. In the most extreme case, that of critical thinking as an A Level, students are required to process huge amounts of knowledge, only for students to be tested on skill rather than knowledge. Critical thinking is a transferable skill, because it is useful anywhere there is an argument, debate or problem to understand. So, critical thinking can be measured in contexts that students and professionals are already familiar with, in order to ensure they are being tested on critical thinking as opposed to knowledge

about an irrelevant scenario. In the Delphi Report, this is termed “content validity” and it is considered an essential part of the assessment of critical thinking (Facione, 1990). Furthermore, combining critical thinking with relevant subjects can reduce the workload of those setting and marking the test, which is especially useful in an educational setting. This has already been attempted in some areas of education. For example, the Psychological Critical Thinking Exam (PTCE) was designed for a psychology class and became so widely used that it was revised for standardised use in psychology education (Lawson, 1999). However, the specificity of this test makes it difficult to transfer to other subjects. An increased focus on critical thinking within the nursing profession also led to interest in critical thinking assessments from the turn of the century (Rane-Szostak and Robertson, 1996; Staib, 2003).

At Endoxa Learning, critical thinking assessment is integrated within content that is relevant to students. We produce lessons on debates that students already need to learn as part of the curriculum, rather than producing generic scenarios as the basis of the assessment. For example, Figure 4 is based on the theory of plate tectonics, which students learn in KS3 geography. This expands on the “content validity” section of the Delphi Report (Facione, 1990). Students are tested on critical thinking skills without having to learn unhelpful information. Moreover, they can also learn crucial subject knowledge, using a system that stays the same whether they are learning about coastal management, economics or the existence of God.

Conclusion

Endoxa Learning solves three important limitations of assessing critical thinking in education:

1. Information is dual coded to reduce verbal loading.
2. We test critical thinking using content that students already need to learn.
3. Students can be tested on arguments of a range of complexities.

Using this approach, we take the positive aspects of in depth, essay-based assessments and combine them with the lower verbal loading of short form tests.

Given the value of critical thinking tests for employers, we also plan to adapt Endoxa Learning to provide employment aptitude tests. These will share the benefits of the tests discussed in previous sections, while making them even more useful for employers who need a clear measure of critical thinking skills among applicants.

Rosie Brown *Content Creation Leader*

Julian Plumley *CEO*

Endoxa Learning Ltd.

11.05.2022

[Contact us](#) to tell us your thoughts and ideas or [click here](#) to read more articles.

Notes

"Critical thinking is reasonable reflective thinking focused on deciding what to believe or do" (Ennis, 1996, p. 166).

References

Cambridge Assessment (undated) *Critical Thinking Adds Value*. <https://www.cambridgeassessment.org.uk/Images/126128-critical-thinking-factsheet-3.pdf> (Accessed 4th May 2022).

Cambridge International (2019) *Cambridge International AS & A Level Thinking Skills: Mark Scheme*. Cambridge: Cambridge University Press and Assessment.

Ennis, R.H. (1996) Critical Thinking Dispositions: Their Nature and Assessability, *Informal Logic* 18:2&3 pp. 165-182.

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

Facione, P.A. (1990) *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*. Millbrae: California Academic Press.

Feldman, J. (2021) Watson Glaser Critical Thinking Test, *Prospects*. Available at <https://www.prospects.ac.uk/careers-advice/interview-tips/watson-glaser-critical-thinking-test#which-employers-use-the-watson-glaser-test> (Accessed 26th April 2022).

Pearson (2002) *Watson – Glaser Critical Thinking Appraisal – UK Edition: Practice Test*. Available at: http://www.pearsonvue.com/phnro/wg_practice.pdf (Accessed 4th May 2022).

Rane-Szostak, D., Robertson, J.F. (1996) Issues in Measuring Critical Thinking: Meeting the Challenge, *Journal of Nursing Education* 35:1 pp. 5-11.

Staib, S. (2003) Teaching and Measuring Critical Thinking, *Journal of Nursing Education* 42:11 pp. 498-508.

World Economic Forum (2020) These are the top 10 job skills of tomorrow – and how long it takes to learn them, *World Economic Forum*. Available at <https://www.weforum.org/agenda/2020/10/top-10-work-skills-of-tomorrow-how-long-it-takes-to-learn-them/> (Accessed 26th April 2022).