Online Supplement for

Make Critical Thinking Skills Training Explicit, Engaging, and Effective through Live Debates on Current Political Issues: A Pilot Pedagogical Experiment

By Yu Tao and Ed Griffith

Published in *PS: Political Science & Politics*

To reference the information in this supplement, simply cite the original article and use the number at the bottom of each page (i.e. S1, S2-S5, etc.). The authors can be reached at yu.tao[AT]uwa.edu.au and egriffith[AT]uclan.ac.uk.

Contents

Section A: Annotated Extracts from Live Debates

Extracts from the Live Debate for Intervention 2

Below, we present an extract of this exchange, along with some indicative notes on certain specific aspects of critical thinking skills that we aimed to demonstrate to the students. In this particular exchange, the proposition was 'Trump's decision to withdraw from the Trans-pacific Partnership (TPP) will have a detrimental effect on regional economic integration'.

> Proposer: The TPP would have provided a sound legal and normative basis for improved regional integration.
>
> Opposer: How are you defining 'regional integration' in this statement? [*Here we sought to demonstrate that questioning the definition of a concept is fundamental to the process of deconstructing an argument.*]
>
> Proposer: A process of developing a rules-based system that promotes deeper economic linkages between countries, enhancing development for all.
>
> Opposer: Let's assume that this is true within the twelve countries that originally signed the treaty. What about those that are not included? [*Here we sought to expose the assumptions that lay behind the original statement from the proposer, demonstrating that the revelation of excluded information from an argument can fundamentally undermine its premise.*]
>
> Proposer: The twelve countries already represented a significant proportion of economic activity across the Asia Pacific but, more importantly, would have provided the foundation on which to build a more integrated regional economic system that could have included others. [*Here we demonstrated the skill of extrapolation in defending the proposer's position from the previous critique by using the underlying logic of the original argument.*]
>
> Opposer: This rests on the assumption that others are willing to participate in a system determined not by themselves and also that those countries within the system would be willing to allow them to join. [*Here we showed the value of explicitly exposing the assumptions underlying the argument which might otherwise remain hidden and, therefore, never challenged*.]
>
> Proposer: Even so, that does not disprove the contention that the TPP would have promoted regional economic integration or that its removal is detrimental to the process.
>
> Opposer: What it means is that the TPP would have prevented wider regional economic integration. It was a barrier to this because it was exclusionary. Its removal from the regional infrastructure opens up space for a more comprehensive regional integration process driven by China's growing leadership on this issue, as evidenced by its commitment to the 'one belt, one road' initiative. [*Here we*

demonstrated the importance of building on the previous points made to strengthen one's case and of illustrating the points with additional, relevant, information.]

Extracts from the Live Debate for Intervention 3

As shown in the following exchange extracted from the record of this discussion, we demonstrated to the students that one can concur with an argument through applying essential critical thinking skills such as identifying and elaborating the fundamental logic underlying a narrative.

> Proposer: While there are many issues that divide the region, one of the greatest threats that every single state in the region faces is climate change. Furthermore, it is an issue that by its very nature requires cooperation.
>
> Seconder: I agree. To elaborate, the key point that you have identified is that the boundaries that divide these states are artificially created. The challenge of climate change, however, does not respect lines drawn on a map. [*Here we demonstrated again the fundamental skill of identifying and exposing assumptions. However, on this occasion we showed that this skill does not necessarily need to be used to highlight a potential weakness; it can also be used to identify the strength of an argument.]*

Section B: Examples of SQEs

SQE Example 1

Provide a critical review of EITHER chapter 6 OR chapter 8 of Bruce Cumings' book *Korea's Place in the Sun*.

Source:
Cumings, Bruce. 2005. *Korea's Place in the Sun: A Modern History*. New York: W.W. Norton.

SQE Example 2

Critically assess Stubbs' analysis of ASEAN's ability to lead the regionalisation process in the Asia Pacific.

Source:
Stubbs, Richard. 2014. "ASEAN's leadership in East Asian region-building: strength in weakness." *The Pacific Review* 27(4): 523-541.

OR

Critically assess Dent's view of the prospects for East Asia's energy diplomacy.

Source:
Dent, Christopher M. 2013. "Understanding the energy diplomacies of East Asian states." *Modern Asian Studies* 47(3): 935-967.

Section C: Marking Rubric (and Considerations underlying the Rubric)

The question of what it means to think critically has been widely explored and examined (Almeida *et al* 2011b; Cuccio-Shirripa and Steiner 2001; Ennis 1969; 1996; 2004; Meyer 1994). It has been long seen as a staple of the social sciences, providing one of their *raisons d'etre* in the face of questions over their value compared with STEM subjects (Almeida *et al* 2011a). Whilst it is beyond the remit of our article to dissect fully the discussions around the basis of critical thinking that stretch back to ancient Greece, we list here some considerations that underpin the marking rubric.

As many academics have posited, there is a direct link between critical thinking and the act of questioning knowledge bases (Cuccio-Shirripa and Steiner 2001). Browne and Freeman (2000) even see the starting point of critical thinking as being a series of questions that seek to expose the structures of an argument. Such questions include evaluation of the evidence provided in terms of quality and reliability, but also seek to assess the argument's persuasiveness and to explore other reasonable conclusions that could be drawn. Therefore, we consider the most fundamental skill for critical thinking is about questioning the assumptions that underpin an argument and exploring the relevance or reliability of the sources of information provided.

Critical thinking also requires a skill that Yalom (1980, 312) described as "simultaneous ambivalence", the ability to be clearly focused on the for and against in any given argument. More explicitly, Johnson and Blair (2006, 50-51) describe such skill as "to admit in principle the possibility that your premises do not constitute good grounds for your conclusion (even though at the moment you think they do)". That is to say, critical thinking is not merely about challenging the premises of an argument for the sake of it, but of questioning all reasonable approaches to the facts in hand in order to ascertain the most convincing explanation.

Guided by these essential principles of critical thinking, we developed the following marking rubric to measure four skills that are widely identified by relevant literature as the most essential to critical thinking (e.g. Cottrell 2017, Roy and Macchiette 2005, Johnson and Blair 2006, Cuccio-Schirripa and Steiner 2000).

Table C1 Marking Rubric of the SQEs

	Quality of Argument	Depth of Analysis	Use of Evidence
89-96 Exceptional First Class	Directly addresses the implications and assumptions in a challenging and authoritative way.	Exceptional analysis with comprehensive arguments and authoritative consideration of wider implications.	Exceptionally convincing conclusions well-supported by the relevant evidence.
74-81 First Class	Directly addresses the implications and assumptions in a sophisticated way.	Excellent analysis with comprehensive arguments and appropriate consideration of wider implications.	Highly convincing conclusions well-supported by the relevant evidence.
62-68 Upper Second Class	Directly addresses the implications and assumptions.	Analysis is thoughtful, clear and ordered.	Convincing conclusions supported by the relevant evidence.
52-58 Lower Second Class	Largely addresses the implications and assumptions but may be less focused in some areas.	Some evidence of analysis but a tendency toward description may be evident and ideas may be expressed only in broad terms.	Evidence is presented but it may not have been engaged with critically.
42-48 Pass	Does not consistently address the implications and assumptions.	Largely descriptive with limited analysis.	An adequate understanding of a limited range of material.
25-35 Fail	May be incomplete or irrelevant.	Over-dependent on description with little or no indication that key issues have been understood.	May not go beyond superficial paraphrasing.
10 Insubstantial Attempt	Not relevant.	Inadequate description. No analysis.	No supporting evidence provided.
0 No Attempt	Non-submission.	Non-submission.	Non-submission.

Notes: The marking bands are discrete because the university has sought to avoid giving student 'ambiguous scores' that are at the border of each level. For example, in the UK system, normally 70 is the threshold for a 'first-class' grade as opposed to an 'upper second-class' performance. To highlight the significantly different expectations between a 'first-class' and an 'upper second-class', the university requires all academics to score 74 for the lowest possible 'first-class' grade and 68 for the highest 'upper second-class' performance.

Section D: Additional Information on the SQE Results in the 2016/17 Academic Year

The students' final grade for this piece of assessment was an average of their four highest grades. Although not every student completed all five short-answer question exercises, the overall participation rate was high, with 80% of students (36 out of 45) completing all the exercises. Of the remaining nine students, five completed four exercises and hence met the minimum requirement of participation for this assessment. The other four students completed either two or three exercises. Although these four students failed to generate a score for this assessment, all their submitted works were marked at the same time, and in the same way, as those submitted by the other students. As a result, the dynamics of their performance in the short-answer question exercises they attempted also reflect the effects of our experiment. Hence, we also included the scores of their completed exercises in the dataset.

Figure C1 presents a box-dot plots chart that offers a straightforward impression on the effects of our interventions. In this chart, each dot represents the score that a student received in an SQE. The depth of each box represents the inter quartile range of the overall performance of the class in each SQE, and the line in the middle of the boxes represents the median score.

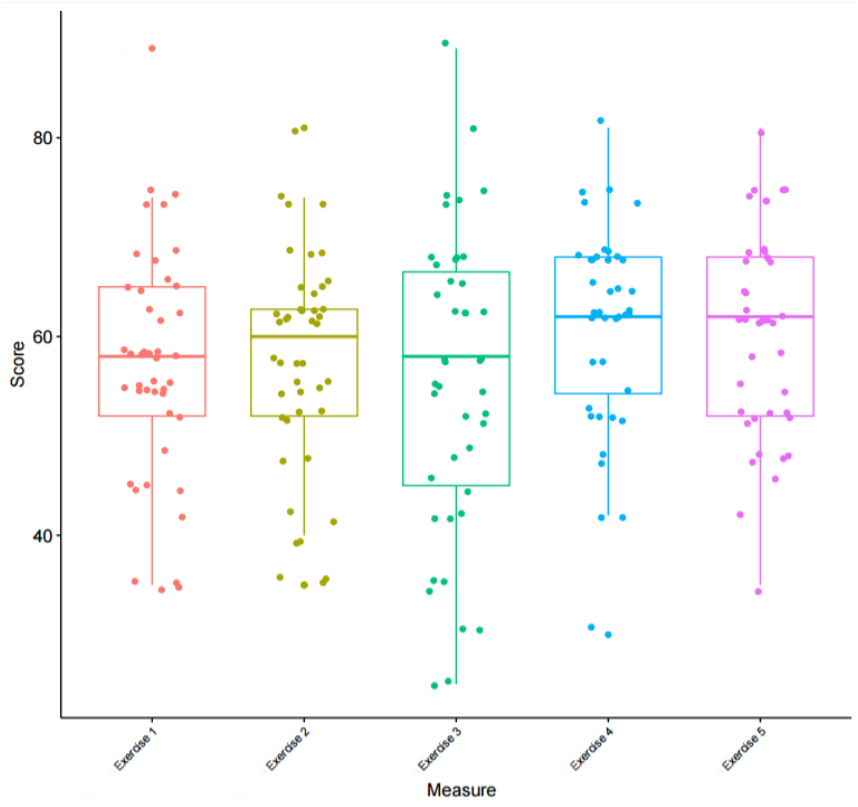

Figure D1 Box-dot Plots of Short Questions Exercise Results[1]

[1] We thank Ye Wang for producing this graph.

Further to the straightforward illustration presented in Figure C1, Table C2 presents the descriptive statistics of the five SQEs in 2016/17, with the timing of all three interventions clearly identified. Confirming the findings we interpreted from Figure C1 (see the main paper), the statistical results presented in Table C2 also show that the students' performance improved significantly after being fully exposed to our explicit demonstration on critical thinking skills through issue-based live debates, with their average scores rising over 60 and their median scores unprecedentedly reaching 62 in SQE4 and SQE5.

Table D2 Descriptive Statistics of the SQE Results

	SQE1	**SQE2**		**SQE3**		**SQE4**		**SQE5**
Time	Week 4 Thursday	Week 6 Tuesday	Intervention 1 (Week 7 Thursday)	Week 8 Tuesday	Intervention 2 (Week 9 Thursday)	Week 10 Tuesday	Intervention 3 (Week 10 Thursday)	Week 12 Thursday
Participants	43	44		43		40		41
Average Score	57.05	57.48		55.44		61.20		60.32
Highest Score	89	81		89		81		81
Lowest Score	35	35		25		48		35
Median Score	58	60		58		62		62
Standard Deviation	11.73	10.88		15.22		10.21		10.34

Putting aside the less quality data of SQE3, we can clearly see that a significant distinction exists in the students' overall performance between their first two short-answer questions exercises (both before the interventions) and their final two (both after their substantial exposure to the interventions). On average, the score of our students increased almost 3.5 points from 57.27 (the arithmetic mean of the average scores of SQE1 and SQE2) to 60.76 (the arithmetic mean of the average scores of SQE4 and SQE5). This is a remarkable improvement in the context of the conventional British system of grading, where it is relatively rare for students to get scores lower than 40 (fail) or higher than 70 (first-class/distinction). Even taking all the 'outlier' scores in our dataset into calculations (with 25 as the lowest mark and 89 as the highest mark - both are indeed very extreme cases), the average performance of all students participating in our pedagogical experiment increased by more than 5% of the overall score range of 64 after we explicitly demonstrated and debriefed a wide range of critical thinking skills through issue-based live debates.

Section E: Robustness Check of the SQE Results in the 2016/17 Academic Year

As shown in the following tables, the pattern of the dynamics of the students' performance in different SQEs appear to be mostly similar among these subgroups, and between them and the whole 2016/17 cohort, suggesting the findings reported in Table 2 are robust.

For all tables in this section, in each non-header grid, the number in the first line displays the paired differences (which is equal to the mean score of the earlier short question exercise subtracted from the mean score of the latter short question exercise, e.g. SQE2-SQE1), the bracketed number in the second line displays t value, and the N number in the third line displays the number of pairs included in a particular t-test. The level of statistical significance is shown by asterisks, where * indicates $p<0.1$ and ** indicates $p<0.05$.

Table 2 Paired T-test Results (the 2016/17 cohort)

	SQE1	SQE2	SQE3	SQE4
SQE2	0.651 (0.380) N=43			
SQE3	-0.756 (-0.429) N=41	-1.762 (-0.785) N=42		
SQE4	3.053* (1.971) N=38	4.179** (2.312) N=39	5.103** (2.557) N=39	
SQE5	2.462* (1.986) N=39	3.150* (1.780) N=40	3.800** (2.321) N=40	-0.846 (-0.616) N=39

Table E1 Paired T-test Results (female students, the 2016/17 cohort)

	SQE1	SQE2	SQE3	SQE4
SQE2	0.880 (0.379) N=25			
SQE3	-2.125 (-1.409) N=24	-2.200 (-0.813) N=25		
SQE4	3.682* (1.823) N=22	7.304** (4.172) N=23	4.272* (2.137) N=22	
SQE5	2.261* (1.761) N=23	4.041** (2.113) N=24	2.565** (1.395) N=23	-1.261 (-0.820) N=23

Table E2 Paired T-test Results (male students, the 2016/17 cohort)

	SQE1	SQE2	SQE3	SQE4
SQE2	0.478 (0.104) N=18			
SQE3	-4.824 (-1.640) N=17	-0.588 (-0.171) N=17		
SQE4	3.218* (1.883) N=16	1.312* (0.924) N=16	5.470** (2.190) N=17	
SQE5	2.312* (1.612) N=16	1. 846* (1.780) N=16	6.823** (2.298) N=17	-0.250 (-0.098) N=16

Table E3 Paired T-test Results (domestic students, the 2016/17 cohort)

	SQE1	SQE2	SQE3	SQE4
SQE2	0.565 (0.285) N=23			
SQE3	-0.954 (-0.456) N=22	-1.681 (-0.712) N=22		
SQE4	3.714* (2.165) N=21	3.667** (2.103) N=21	1.842* (1.543) N=21	
SQE5	2.142* (1.977) N=21	3.238* (1.746) N=21	1.823* (1.134) N=21	-0.750 (-0.456) N=20

Table E4 Paired T-test Results (international students, the 2016/17 cohort)

	SQE1	SQE2	SQE3	SQE4
SQE2	0.725 (0.681) N=20			
SQE3	-0.520 (-0.356) N=19	-2.177 (-0.956) N=20		
SQE4	2.832* (1.225) N=17	5.778** (3.407) N=18	9.889** (2.839) N=18	
SQE5	2.833* (1.911) N=18	3.520* (1.543) N=19	7.736** (2.989) N=19	-0.947 (-0.417) N=19

Section F: Robustness Check of the SQE Results in the 2017/18 Academic Year

As shown in the following tables, the pattern of the dynamics of the students' performance in different SQEs appear to be mostly similar among these subgroups, and between them and the whole 2017/18 cohort, suggesting the findings reported in Table 3 are robust.

For all tables in this section, in each non-header grid, the number in the first line displays the paired differences (which is equal to the mean score of the earlier short question exercise subtracted from the mean score of the latter short question exercise, e.g. SQE2-SQE1), the bracketed number in the second line displays t value, and the N number in the third line displays the number of pairs included in a particular t-test. The level of statistical significance is shown by asterisks, where * indicates p<0.1 and ** indicates p<0.05.

Table 3 Paired T-test Results (the 2017/18 cohort) [As in the main paper]

	SQE1	SQE2	SQE3
SQE2	0.809 (0.689) N=68		
SQE3	-7.701** (-4.910) N=67	-8.191** (-5.462) N=68	
SQE4	-0.894 (-0.599) N=66	-1.373 (-0.964) N=67	6.652** (3.795) N=65

Table F1 Paired T-test Results (female students, the 2017/18 cohort)

	SQE1	SQE2	SQE3
SQE2	1.667 (0.970) N=36		
SQE3	-4.829** (-2.132) N=35	-6.429** (-3.268) N=35	
SQE4	0.714 (0.295) N=35	-1.200 (-0.545) N=35	4.911* (2.167) N=36

Table F2 Paired T-test Results (male students, the 2017/18 cohort)

	SQE1	SQE2	SQE3
SQE2	-0.156 (-0.098) N=32		
SQE3	-10.844** (-5.286) N=35	-10.061** (-4.442) N=33	
SQE4	-2.710* (-1.702) N=31	-1.563 (-0.868) N=32	8.500** (3.154) N=32

Table F3 Paired T-test Results (domestic students, the 2017/18 cohort)

	SQE1	SQE2	SQE3
SQE2	2.315 (1.806) N=38		
SQE3	-8.595** (-4.202) N=37	-10.865** (-5.328) N=37	
SQE4	-0.583 (-0.374) N=36	-2.861 (-1.682) N=36	7.800** (3.119) N=35

Table F4 Paired T-test Results (international students, the 2017/18 cohort)

	SQE1	SQE2	SQE3
SQE2	-1.100 (-0.527) N=30		
SQE3	-6.600* (-2.688) N=30	-5.000* (-2.373) N=31	
SQE4	-1.267 (-0.464) N=30	0.355 (0.151) N=30	5.354* (2.174) N=30

Bibliography

Almeida, Patrícia Albergaria, Teixeira-Dias, José Joaquim, & Martinho, Chinthaka Damith Balasooriya. 2011a. "The Interplay between Students' Perception of Context and Approaches to Learning." *Research Papers in Education* 26 (2): 149-169.

Almeida, Patrícia Albergaria, Pedrosa-de-Jesus, Helena, & Watts, D. 2011b. "Kolb's Learning Styles and Approaches to Learning through the Use of Students' Critical Questions." In Rayner, Steve & Cools, Eva (eds), *International Perspectives on Style and Difference in Human Performance*. London: Routledge.

Browne, Neil and Freeman, Kari. 2000. "Distinguishing Features of Critical Thinking Classrooms". *Teaching in Higher Education* 5 (3): 301-309.

Cottrell, Stella. 2017. *Critical Thinking Skills: Effective Analysis, Argument and Reflection*. London: Palgrave Macmillan.

Cuccio-Schirripa, Santine, and H. Edwin Steiner. 2000. "Enhancement and Analysis of Science Question Level for Middle School Students." *Journal of Research in Science Teaching* 37(2): 210-224.

Ennis, Robert. 1969. *Logic in Teaching*. Englewood Cliffs, NJ: Prentice-Hall.

Ennis, Robert. 1996. "Critical Thinking Dispositions: Their Nature and Accessibility." *Informal Logic* 18 (2-3): 165-182.

Ennis, Robert. 2004. "Applying Soundness Standards to Qualified Reasoning." *Informal Logic* 24 (1): 23-39.

Johnson, Ralph H. and J. Antony Blair. 2006. *Logical Self Defense*. New York: International Debate Education Association.

Meyer, Michael. 1994. *Rhetoric, Language and Reason*. Pennsylvania: The Pennsylvania State University Press.

Roy, Abhijit, and Bart Macchiette. 2005. "Debating the Issues: A Tool for Augmenting Critical Thinking Skills of Marketing Students." *Journal of Marketing Education* 27 (3):264-276.

Yalom, Irvin. 1980. *Existential Psychotherapy*. New York: Basic Books.