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

Abstract

We piloted a pedagogical experiment to find out whether students can benefit from explicit demonstrations on critical thinking skills through live debates between two instructors on current political issues that are relevant to, but not necessarily a specific part of, the curriculum. The empirical results show that through a series of interventions in the form of explicit demonstrations and debriefs on critical thinking skills over these issue-based live debates, the students' academic performance can significantly improve over a relatively short period of time. This result, we suggest, demonstrates that training the students' critical skills through explicit, engaging pedagogy is not only economical in practical and pragmatic terms, but also proven to have at least significant immediate, short-term effects in a setting where there is a high proportion of first generation undergraduate students of varying abilities and backgrounds.

Authors

Yu Tao & Ed Griffith

Yu Tao is a Lecturer in the School of Social Sciences at the University of Western Australia. He can be reached at yu.tao@uwa.edu.au. Ed Griffith is a principal lecturer in the School of Language and Global Studies at the University of Central Lancashire. He can be reached at egriffith@uclan.ac.uk.

Acknowledgement

We are very grateful for the inspiring and helpful comments from the editor and reviewers, as well as the kind support from all students who participated in this pedagogical experiment. All remaining errors are our own.

Critical thinking, according to the well-accepted succinct definition of Ennis (1985, 45), is 'reflective and reasonable thinking … focused on deciding what to believe or do'. The complex process of critical thinking involves a wide range of skills. Among these the most essential ones include: identifying the logistic structure of an argument; assessing whether a claim is made on sound empirical grounds; weighing up opposing arguments and evidence fairly; and seeing behind the surface level and through false assumptions (Cottrell 2017, 2). These skills are vital for students to make sense of important issues in and beyond the

discipline of political science (Atwater 1991, Cohen 1993). Providing students with an intellectual tool kit of critical thinking has been widely accepted as an essential function of modern higher education (Hanscomb 2015). Moreover, the abilities to deconstruct a narrative, to question the assumptions that underpin a claim, to explore the relevance and reliability of the sources of information provided, and to appreciate the logic and reasons behind an argument different from one's own are crucial for responsible citizens to engage in politics (Lamy 2007; ten Dam and Volman 2004). This is particularly so in facing the recent resurgence of populism, racism, and hate discourses.

Yet for many educators, including ourselves, developing students' critical thinking skills is a challenging task (Çavdar and Doe 2012). It is sometimes assumed that students will somehow 'absorb' the skills of critical thinking through 'immersing' themselves in the environment of higher education, through observing their peers, or through reading the literature associated with their degree programs (Ennis 1989). However, having instructed in a wide range of higher education settings, we observed that not all students are able to pick up critical thinking skills through their normal university experiences and class participation. This observation, combined with feedback we frequently received from students regarding the difficulties they had with grasping the fundamental tenets of critical thinking, motivated us to look beyond the conventional 'immersion' approach and seek strategies that are more explicit and effective in helping students develop their critical thinking skills.

Existing research suggests that issue-based live debates are effective in explicitly demonstrating some of the most essential critical thinking skills (Roy and Macchiette 2005). Pedagogical experiments have shown that a 'crossfire-style' live debate between two instructors performed in front of a class can effectively heighten students' interest and engagement in the academic discipline of political science, and such a performance can also demonstrate the feasibility of disagreement or critique in a civil manner, dispelling a common

misperception that political disagreement is necessarily conflictual (Baumgartner and Morris 2015). Inspired by these findings, we piloted a pedagogical experiment on a group of 45 final-year undergraduates taking a class on politics and international development in East Asia. During the experiment, we performed a regular section of issue-based live debates between ourselves during the weekly lectures and explicitly debriefed the critical thinking skills employed during our debates. We also assessed the students' critical thinking skills through a series of standardised short-answer question exercises (SQEs), which formed part of the students' summative assessment for the course, before and after the interventions. The empirical results demonstrate a positive correlation between our experimental interventions and our students' performance in the SQEs designed to test their critical thinking skills. This suggests that live debates on current political issues, accompanied with immediate explicit debriefs and articulations on the critical thinking skills used, are indeed effective in improving the students' critical thinking skills – at least in the short term and in certain higher education settings.

INTERVENTIONS

We conducted our pedagogical experiment during a twelve-week final-year undergraduate course titled 'Development and Change in the Asia Pacific' during the 2016/17 academic year. This course is designed to deepen the students' understanding of the processes of political and economic development in the Asia Pacific region, with a particular focus on China, Japan and Korea. In addition to the subject-specific knowledge, critical thinking skills are also among the course's learning outcomes, as is commonly the case in British universities. This semester-long course had two two-hour sessions in each teaching week, and all students were taught together in the same group.

Existing research has posited a direct link between critical thinking skills and the act of questioning knowledge bases (Cuccio-Shirripa and Steiner 2000). Live debates, in this regard, are effective tools to teach critical thinking skills because they create arenas in which participants have to apply a variety of critical thinking skills to question the premises of opposing arguments and to ascertain the most convincing explanation. Moreover, training critical thinking skills in political science requires educators to 'bring students into contact with the world outside their own unchallenged perceptions of it' (Hoefler 1994), and live debates on current political affairs can vividly demonstrate to the students the necessity of admitting 'in principle that the possibility that one's premises do not always constitute good grounds for one's conclusion' (Johnson and Blair 2006, 50-51).

To demonstrate how to apply critical thinking skills, in late 2016, we intervened in the normal teaching and learning activities of our course with a regular section of live debates between us on current political issues. Each of our intervention sessions lasted for approximately fifteen minutes, which comprised (1) a brief introduction during which we identified the topic for the session, clarified the rules including how we take sides in the live debate, and explicitly reminded our students that the main purpose of our live debate was to demonstrate the critical thinking skills that were to be evaluated through formal assessments, (2) a live debate during which we questioned, critiqued, or critically concurred with each other's ideas, and (3) a short yet clear after-debate debrief during which we explicitly commented on the lessons (and sometimes the mistakes) from our application of critical thinking skills during our debates.

Our live debates, each took approximately ten minutes, focused on current political issues that were relevant to, but not specifically a part of, the curriculum. For example, in November 2016, we focussed our second debate on the United States' withdrawal from the Trans-Pacific Partnership shortly after the then President-elect Donald Trump announced that

he would honour the promise he made to do so during the election campaign.[1] Prior to the debate, we briefly discussed the possible ramifications of this action. When the lecture started, we flipped a coin in front of the class to decide which position each of us would take in the debate. We did this deliberately, with the hope to demonstrate explicitly to the students that critical thinking skills are needed and helpful regardless of one's position in an academic argument or debate. This intention, along with an idea of the skills that we would like students to observe during the debates, was clearly communicated to them prior to the actual debates.

During our debates, we made efforts to demonstrate a variety of critical thinking skills that are widely identified as essential for students in and beyond the discipline of political science. These included questioning the definitions of terms, identifying pertinent ideas and factors, reasoning, adaptation to context, and, in particular, distinguishing opinions from facts (Fitzgerald and Baird 2011). In addition, from previous teaching experience we were conscious that some students may confuse critical thinking with criticism. To demonstrate that critical thinking skills can, and should, be applied to deepen and enrich the discussions in which the participants fundamentally agree, in our final discussion we deliberately chose to take the same side on the proposition, which posited that the issue of climate change presents an opportunity for the Asia-Pacific region to deepen international cooperation.

Our scepticism on the assumption that students can somehow 'naturally' grasp critical thinking skills by immersing themselves in the environment of higher education led us to make targeted efforts to articulate explicitly what are critical thinking skills and how one can apply them. To ensure that our students were completely conscious of what we were trying to teach them through the live debates, after each debate we always spent a few minutes elaborating the lessons – and sometimes the mistakes – from our application of critical

thinking skills. Students were also invited to participate in these debriefs through asking questions and offering comments on the critical thinking skills we employed during the debates.

MEASURES

Altogether, we introduced three interventions (live debates) during the experiment period. To measure the effectiveness of these interventions, we introduced a series of five SQEs as a component of the formal assessment for the class. These SQEs were spaced out across the semester at two-week intervals. Each SQE gave the students a choice of two academic articles or book chapters to assess critically.[2] The students were required to write no more than 200 words articulating why they agree, disagree, or partially agree with the main argument presented in the selected text.

The students were informed that there were no 'right' or 'wrong' answer to the questions and that their grade depended only on the level of competence they displayed in applying critical thinking skills to the tasks set. Furthermore, it was made clear to them that they were expected to learn these skills from observing the live debates, listening to our introductions, and participating actively in the debrief sessions. Following the standard procedure for summative assessments at the university in which the experiment was conducted, all answers were marked anonymously by a main examiner who followed a grading rubric that specifically focussed on critical thinking skills.[3] For each SQE, a random sample of answers in each grade band was independently second marked, following the same rubric used by the main examiner. The university procedure requires that any disputed cases should be discussed between the two examiners and, when the first examiner is successfully challenged during such a discussion, the answers would be re-marked in their entirety. In the

year in which we conducted this experiment, no such action was necessary. Finally, at the end of the semester, an external examiner from another university also randomly selects several answers in each marking band of all SQEs to review the grades in the context of the rubric, and to benchmark them against the relevant national academic quality assurance framework. In the particular year when we conducted our experiment, the external examiner was not only satisfied with the marks but also praised the quality and consistency of the marking process.

< Table 1 is about here. >

We outline the experiment sequence in Table 1. After an initial period for introducing the course and going over some basic knowledge regarding critical thinking skills, we introduced the first SQE in Week 4 to obtain the baseline information regarding the critical thinking skills of our students. As a pilot experiment, we did not separate our students into a treatment group and a control group. To mitigate this, we did not introduce any intervention between the first two SQEs, so that a comparison between the results of these two SQEs could enable us to identify the 'normal' trend of academic performance when the students are exposed to ordinary teaching and learning sessions. We introduced our first intervention shortly before SQE3, and we took opposite positions in that debate. A similar intervention, during which we once again took opposite positions, was introduced between SQE3 and SQE4. Our final intervention was conducted between SQE4 and SQE5, and on this occasion we deliberately chose to concur with each other.

RESULTS

The empirical results of the SQEs show that our pilot experiment was a success, suggesting that demonstrating critical thinking skills explicitly through live debates on current political issues can indeed significantly improved these skills in our students. Generally speaking, the

overall performance of the class in SQE4 and SQE5 was noticeably better than it had been in the previous three. This upturn followed our second and third interventions. The result of SQE3 stands out as having, by far, the highest diversity of scores. Whilst the medium score of SQE3 was similar to that of SQE1 and even slightly lower than that of SQE2, its higher quartile is noticeably higher than those of both previous measures, suggesting at least some students started grasping the critical thinking skills that we hoped to teach them immediately after the first intervention.[4]

Considering that the aggregated scores may be affected by the presence or absence of certain students, we further examined the impacts of our interventions on the individual level by conducting a series of paired T-tests to compare each student's performance in different SQEs. As shown in Table 2, whilst, on average, many students performed slightly better in SQE2 and slightly worse in SQE3, the difference in their performance during the first three SQEs is not statistically significant. However, after being exposed to at least one purposely-designed issue-based live debate in the full circle of preparing for their assignment, on average each student scored 3 to 5 points (or between 4.7% and 7.8%) higher in SQE4 than in the previous three exercises, and these results are statistically significant. The results of SQE5 followed the same pattern, confirming that the performance of students significantly improved after we explicitly demonstrated to them essential critical thinking skills through live debates based on current political issues.[5]

< Table 2 is about here. >

We also performed paired T-tests in the subgroups of male, female, domestic, and international students. 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 sample, suggesting the findings reported in Table 2 are robust.[6]

The empirical results reported in Table 2 also show that there is no linear progression in the students' performance through SQE1 to SQE5 – their performance improved in SQE2 and SQE4, but decreased in SQE3 and SQE5, despite the general trend of improvement during our experiment. It is therefore fair to accept that the improvement in the students' performance cannot be simply explained as being a result of their increased familiarity with the task or the topics of the course.

To further check the robustness of our results, we also examined the SQE results of the same course taught in the 2017/18 academic year. Although the requirements and marking processes for the SQEs are identical between the two academic years, we were not able to perform issue-based live debates in 2017/18 because one of us relocated to another country and their replacement was not appointed when the course was taught. The student cohort of 2017/18 is about 50% larger than that of 2016/17, but the two cohorts are otherwise generally similar. Therefore, though not a deliberate design, the 2017/18 cohort serves as a decent *de facto* control group in our pilot experiment.

< Table 3 is about here. >

As demonstrated in Table 3, when intervention is not preformed, the students' performance in SQEs did not naturally increase over time. Apart from the significantly worse result of SQE3, there is no significant difference between the students' performance in the other SQEs.[7] Our robustness test further confirms that such a pattern also exists in the subgroups of female, male, domestic, and international student.[8] These results not only enhance our confidence in believing that the improvement of the students' performance in the 2016/17 academic year was indeed a consequence of the interventions, but also vividly demonstrate that simply 'immersing' students in the normal teaching and learning activities in university does not automatically lead to the development of their critical thinking skills.

LESSONS

The encouraging results of our pilot pedagogical experiment show that training students in critical thinking skills is an achievable task despite its challenging nature, and that even a modest number of explicit demonstrations on critical thinking skills through purposely-designed live debates on current political issues can have a noticeable immediate positive impact on the students' academic performance.

Our results add to the body of literature that indicates students learn critical thinking skills much more effectively through explicit rather than implicit training (Halpern 1998). Before this experiment, our previous attempts to incorporate critical thinking skills into the curriculum achieved little success. We had selected reading materials that were not only relevant to the curriculum but also exemplary in applying critical thinking skills, but it appeared to be insufficient to assume that the students would 'naturally' pick up the necessary skillset to understand and apply critical thinking through conventional teaching and learning activities such as reading literature and in-class discussions. The contrast between our previous experience and the results of this pilot experiment has led us to believe that it is more efficient to teach critical thinking skills through explicit demonstration than through the conventional 'immersion' or 'infusion' approaches, at least in settings similar to the large, diverse, modern public university like the one in which we conducted the experiment.

Our results further suggest that different strategies of explicitly teaching critical thinking skills may also vary in their effectiveness. We had also previously attempted to be explicit in articulating critical thinking skills to our students through standalone workshops and training sessions, most of which centred on straightforward introductions of the abstract concepts and epistemological foundations of critical thinking skills which were

predominately illustrated through artificial examples. Despite the considerable extra time and energy that we invested in organising these events (which in many cases were not recognised in our workload), it was difficult to secure either a satisfactory turnout rate (when these sessions were made optional) or a decent level of attention and enthusiasm (when these sessions were made compulsory). The level of success achieved through the pilot experiment introduced in this paper, however, far exceeds any progress that we had previously made through other methods. To ensure the students' interest in our live debates remained high, we drew topics from current affairs that had tangible connections to the areas being addressed in class. This proved useful. During the live debates, we could clearly feel that most of the students were enthusiastic and engaged. In the anonymous course evaluation at the end of the semester, several students identified our live debates as the aspect of the class that they enjoyed the most.

Our success was achieved with a moderate amount of recourse. Once the fundamental design of our pedagogy was decided, we spent only about half an hour before each intervention session to go through both the possible scenarios in our upcoming debate and the key critical thinking skills that we would like to cover. We normally did this as a part of our routine casual exchange of ideas during coffee breaks. The fact that we had been working together in the same course team for a couple of years probably helped us reduce the time required for preparation, but in our opinion even a newly-formed course team could easily replicate what we did as long as a healthy working relationship exists between the two instructors co-delivering the live debates.

The relatively modest amount of time and energy we spent in preparing and executing the interventions means that our pedagogy requires low investment in human resource. The effort we made in designing and delivering the issue-based live debates contributed to the general preparation and delivery of our course, and hence did not noticeably increase our

workload. Furthermore, despite the need for the training to be explicit and a period of time designated for its completion, our live debates did not impact on the time spent on the subject matter in class. Our students were able to benefit from witnessing an informed discussion of issues that were relevant to their curriculum (and assessments) whilst simultaneously improving their critical thinking skills.

It is worth emphasizing that one objective we hoped to achieve through our live debates and debrief sessions was to exemplify that critical analysis does not need to be hostile in its nature. This is an essential aspect of the students absorbing the critical thinking skills into their habitual behaviour, making them reasonable and responsible citizens. We believe this objective, though not explicitly measured, was also achieved. This was reflected in comments we received from students, who observed that our debates, while robust and rigorous, were always good natured and ended with us either demonstrating where common ground had been found or accepting the difference that we identified between the philosophical roots of respective viewpoints..

REFLECTIONS

As a pilot project, our experiment was not without shortcomings. For example, although we carefully examined the dynamics of SQE scores in each subgroup defined by students' gender and country of origin, due to the size of our sample we were not able to directly measure whether these personal characteristics actually have significant influence on how our pedagogy affects students on the individual level. In addition, although our students clearly benefited from the purposely-designed interventions in a measurable way, it is not yet clear if this rate of improvement could continue to be delivered if a longer period, or a larger amount, of similar interventions were employed. It should also be mentioned that most of our students

come from nonselective, state-funded secondary schools, and very few of them had previously been exposed to extensive training on critical thinking skills through debates or other engaging forms before this experiment. Unfortunately, we were not able to obtain the data regarding each individual student's socioeconomic background for a specific investigation into this matter. However, it is perhaps legitimate to question whether our pedagogy would generate a similar scale of success when it is applied to those who have been very familiar with, and practiced at, debating. Certainly, further research in this area would be valuable. That said, all these shortcomings generate testable hypotheses for subsequent investigation and experiments, which is itself an objective for pilot experiments.

To conclude, the nature of our pilot experiment was exploratory, and our findings remain encouraging in this regard. Our success came despite a class of varied abilities and our success was achieved with just a few sessions of issue-based live debates. This suggests that our pedagogy could easily be deployed in similar settings for significant benefits, at least in the short term. We hope that the methods and findings reported in this paper offer some insight and inspiration for fellow educators of political science to take on the commonly-faced challenge of developing students' critical thinking skills in higher education.

REFERENCE

Atwater, Thomas. 1991. "Critical Thinking in Basic U.S. Government Classes." *PS: Political Science and Politics* 24 (2): 209-211.

Baumgartner, Jody C., and Jonathan S. Morris. 2015. "The "Crossfire Approach": Attracting Political Science Majors in Large-Lecture Sections of Introductory Courses." *PS: Political Science and Politics* 48 (4): 612-616.

Çavdar, Gamze, and Sue Doe. 2012. "Learning through Writing: Teaching Critical Thinking Skills in Writing Assignments." *PS: Political Science and Politics* 45 (2): 298-306.

Cohen, Mel. 1993. "Making Critical Thinking a Classroom Reality." *PS: Political Science and Politics* 26 (2): 241-244.

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 H. 1985. " A Logical Basis for Measuring Critical Thinking Skills." *Educational Leadership* 43 (2): 44 - 48.

Ennis, Robert H. 1989. "Critical Thinking and Subject Specificity." *Educational Researcher* 18 (3): 4-10.

Fitzgerald, Jennifer, and Vanessa A. Baird. 2011. "Taking a Step Back: Teaching Critical Thinking by Distinguishing Appropriate Types of Evidence." *PS: Political Science and Politics* 44 (3): 619-624.

Halpern, Diane F. 1998. "Teaching Critical Thinking for Transfer across Domains." *American Psychologist* 53 (4): 449-455.

Hanscomb, Stuart. 2015. "Assessment and the Self: Academic Practice and Character Attributes." *Journal of Learning Development in Higher Education* 8: 1-18.

Hoefler, James M. 1994. "Critical Thinking and the Use of Optical Illusions." *PS: Political Science and Politics* 27 (3): 538-545.

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

Lamy, Steven L. 2007. "Challenging Hegemonic Paradigms and Practices: Critical Thinking and Active Learning Strategies for International Relations." *PS: Political Science and Politics* 40 (1): 112-116.

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.

ten Dam, Geert, and Monique Volman. 2004. "Critical Thinking as a Citizenship Competence: Teaching Strategies." *Learning and Instruction* 14 (4): 359-379.

Tables

Table 1 Experiment Arrangements

	Action	Time	Topic	Objectives
The 1st measure	SQE1	Week 4, Thursday		Measuring the baseline level of the student's critical think skills
The 2nd measure	SQE2	Week 6, Tuesday		Confirming the baseline level of the student's critical think skills; measuring whether students' critical thinking skills change without intervention
The 1st intervention	A live debate (we took the opposite sides)	Week 7, Thursday	The effect of the election of Donald Trump on the Asia Pacific	Demonstrating critical thinking skills
The 3rd measure	SQE3	Week 8, Tuesday		Measuring the immediate effect of our methods after the 1st intervention
The 2nd intervention	A live debate (we took the opposite sides)	Week 9, Thursday	The decision of the US to withdraw from the Trans-pacific Partnership	Demonstrating critical thinking skills
The 4th measure	SQE4	Week 10, Tuesday		Measuring the effects of our methods after two interventions
The 3rd intervention	A live discussion (we took the same side)	Week 10, Thursday	The impact of climate change on regional cooperation	Demonstrating critical thinking skills
The 5th measure	SQE5	Week 12, Thursday		Measuring the effects of our methods after three interventions

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

Notes: In each non-header grid, the number in the first line displays the paired differences (which are 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, 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)

	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

Notes: In each non-header grid, the number in the first line displays the paired differences (which are 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, 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 $p<0.05$.

NOTES

[1] An extract of this and the other exchanges mentioned in this paper, along with some indicative notes on certain specific aspects of critical thinking skills that we aimed to demonstrate to the students through these debates, can be found in Section A of the online supplement.
[2] Please refer to Section B of the online supplement for two examples of the SQEs.
[3] Please refer to Section C of the online supplement for the rubric used in the marking.
[4] The students' final grade for this assessment was an average of their four highest grades. There were a few students who joined the course late or withdrew during the semester. However, most students attempted all five SQEs. For more details please refer to the online supplement, where Figure D1 provides a straightforward illustration on the effects of our interventions and Table D2 reports the descriptive statistics of the SQE results (including the number of students attempting each SQE).
[5] On average, students performed slightly worse in SQE5 than in SQE4. The difference, though, is not statistically significant.
[6] Please refer to Section E of the online supplement for results of the robustness test.
[7] Only four SQEs were arranged in 2017/18 due to the university's decision of shortening semesters.
[8] Please refer to Section F of the online supplement for results of the robustness test.

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.

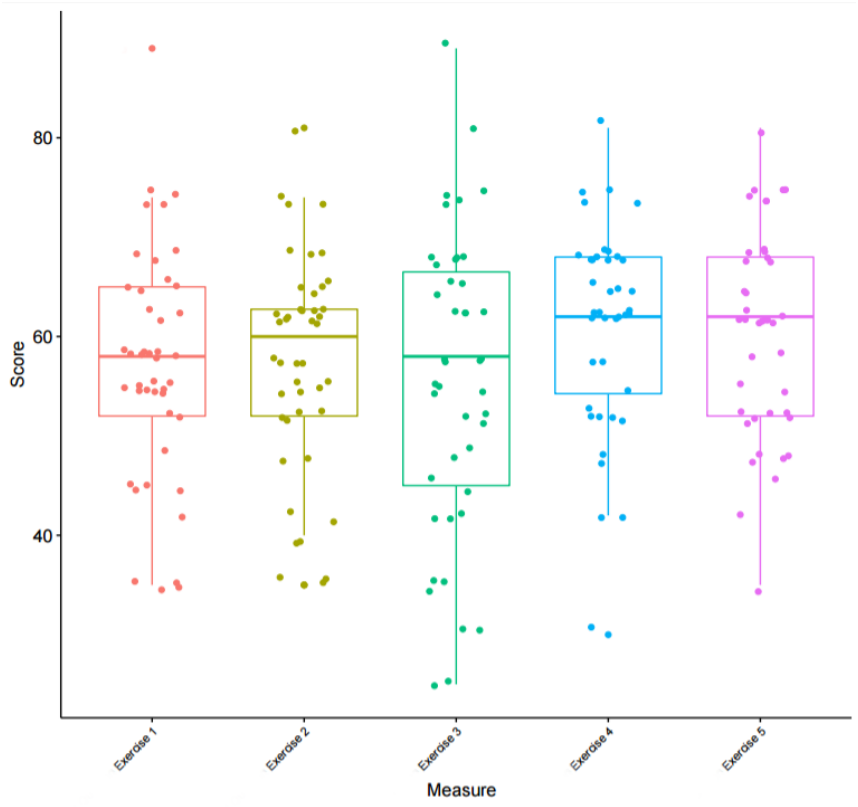

Figure D1 Box-dot Plots of Short Questions Exercise Results[1]

[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.