

Encouraging Critical Thinking in University-Level English Classes in Japan: context, challenges, and practical strategies

Christofer BULLSMITH

Abstract

Critical thinking is a key skill in the modern world. This paper discusses the challenges to encouraging critical thinking in EFL classes in Japan, including the local understanding of key terms, the social and educational context, and the absence of critical thinking in many set curricula and coursebooks. Practical workarounds and detailed suggestions for encouraging critical thinking despite these challenges are provided.

Critical thinking (CT), very briefly and following the consensus definition from the 1990 American Philosophical Association Delphi Report on CT, is ‘purposeful reflective judgement’ about what to believe or do (Facione 2011: 22). What skills are used, what dispositions are necessary, and what actions are taken depend on the context (say, listening to a child describe a monster under their bed, versus reading a war-era history textbook). While this vexes attempts to give a non-trivial characterization of CT in terms of necessary and sufficient conditions, a few examples of uncritical thinking (where someone understands and simply accepts the given content as true) contrasted with a few of critical thinking (where motives, sources, reasoning, and so on are questioned and what is ultimately believed may be radically different to the given content) provide some appreciation of the continuum intended. Frequently present features of CT include questioning, truth-seeking, interpretation, analysis, evaluation, inference, and so on.

CT is a key skill in a fast-changing world, helping people (as moral agents, voters, employees, educators, consumers, and so on) tackle complex and novel problems. The OECD recognizes it as a core ‘21st century skill’ even while employers regularly complain that new graduates lack it.

Challenge one: ‘critical thinking’ is an unpopular label in Japan

Japan’s Ministry of Education, Culture, Sports, Science and Technology (MEXT) website provides a

useful window onto the recent concerns of educators in Japan. It includes not only the official ‘Courses of Study’ (学習指導要領 / *gakushū shidō yōryō*) curriculum standards for all levels of school and all subjects, plus the official comments and guidelines, but also transcripts of the relevant meetings of committees on education, curricula, IT in schools, higher education challenges, and so on.

As a quick orientation to critical thinking in Japan, the entire MEXT site was searched for ‘critical thinking’ and enough was read around each hit to understand the context (what was being said about CT) and/or any interesting related terms (in particular, terms presented as similar to ‘CT’), which were noted for analysis. Typically this meant reading the headings of the text and two to four sentences before and after the sentence of occurrence. This process was repeated for all the related terms found (the first two being the Japanese transliteration and translation of ‘critical thinking’) and iterated for further discovered terms until the terms being examined were no longer directly relevant to CT.

	Term	Occurrences
1	critical thinking	205
2	クリティカルシンキング [<i>kuritikaru shinkingu</i> = critical thinking]	302
3	批判的思考 [<i>hihanteki shikō</i> = critical thinking]	765

In terms of themes, #1 to #3 are in decreasing order of ‘English relatedness’, with the most English-related term (#1) seen largely in discussions of international educational agreements etc. (for example the 2014 Aichi–Nagoya Declaration on Education for Sustainable Development), the Japanese transliteration (#2) seen mostly in discussions of EFL education in Japan but occasionally also in connection with overall goals of education, and the Japanese translation (#3) being used in connection with education overall and with a range of specific areas (particularly EFL, but also research, medicine, and so on).

The main theme seen in connection with #1 is that CT is important internationally; the main theme seen in connection with #2 and #3 is that Japan doesn’t do it (at all, well, or much), but should. Even #3 is still clearly understood as the translation of a foreign idea, something brought in from outside, with regular discussion of why it is particularly difficult in the Japanese educational context. The minutes of meetings make it clear that 批判的 [*hihanteki* = critical] can also mean trigger-happy, hostile, or negative, so #3 (more even than #2) carries a clear connotation of hostility rather than of a constructive, truth-focused attitude or process.

Overall, while senior educators and policymakers in Japan are evidently aware of CT, and be-

lieve that CT should be promoted more in Japan, the discussion seems to be mostly stuck on that CT should be promoted (rather than *how* it should be promoted). Further, most of the discussion is in the context of EFL, suggesting that ‘critical thinking’ and its Japanese translation label a foreign activity which, although attractive to well-travelled senior educators, is ultimately unsuited to the Japanese context.

	Term	Occurrences
4	研究マインド [<i>kenkyū maindo</i> = research mind]	373
5	論理的思考 [<i>ronriteki shikō</i> = logical thinking]	2356

There are, however, terms related to CT that don’t suffer from such negative or foreign connotations. Research mind (#4), for example, occurs mostly in discussion about medical education and the need for a new generation of doctors who can do novel research, but occasionally occurs also in more general discussions of developing the skills required for graduate study and research careers. Again, the main theme is that Japan doesn’t have it, do it, or encourage it well, but ought to. Presumably the skills required for #4 are broadly similar to those required for critical thinking, but the label is apparently far more appealing.

Likewise, logical thinking (#5) lacks the negative connotation of ‘critical’, so educators and policy-makers who shrink from promoting ‘hostile’ thinking in school curricula can agree on the need for logical thinking, and this term appears nearly twice as often as all the CT varieties (#1–#3) combined. The main theme is *how* it is to be encouraged, with a wide range of sub-themes: for example, encouraged in primary school students by teaching basic programming, in graduate students by teaching philosophy and logic, and in various subjects including Japanese, English, and others.

Logical thinking in a broad sense presumably involves the skills listed above as characteristic of CT (questioning, truthseeking, interpretation, analysis, evaluation, and inference). Further, instruction and practice in argument diagramming, a logical thinking skill, has been shown to produce significant gains in critical thinking skills (Twardy 2004), suggesting that CT and logical thinking are indeed closely related.

	Term	Occurrences
6	active learning	164
7	アクティブラーニング [<i>akutibu rāningu</i> = active learning]	5255

The final noteworthy related terms are active learning (#6) and its transliteration (#7). These again lack the negative connotation of CT, and indeed seem to enjoy a positive one. Despite appearing only comparatively recently, they appear 2.3 times as often as logical thinking (#5) and nearly 4.3 times as often as all the CT varieties combined. The main theme, as for #5, is how active learning can be encouraged at all levels of school and in all subjects, with regular discussion of how and why it is challenging in the Japanese context and how overseas educational systems (German, British, American, and so on) encourage it better than the Japanese school system does.

Active learning involves *doing* something while learning, and thinking about what you're doing: the general idea is that writing, discussing, solving problems, and generally engaging in higher-order thinking tasks (evaluating, analyzing, synthesizing etc.) aids understanding and retention (a point likely familiar to EFL specialists as the idea that deeper processing results in better comprehension and retention – see for example Craik & Lockhart 1972). This differs from CT in that it focusses on improving *learning* rather than critical (hostile!) truth-seeking, and therefore sits better with traditional Japanese educational methods. Active learning also likely benefits from being more accessible than CT: encouraging questions, or having students write short summaries and give reasons, thereby keeping them awake and engaged, seems manageable even where the rather more nebulous, abstract, and foreign CT might sound possible only for high-level students with a specially trained teacher.

Overall, then, logical thinking and (even more so) active learning are more common and positive labels in Japan than CT. Discussion of these appears to have moved beyond *that* they should be promoted to *how*, suggesting that individuals and institutions unenthusiastic about CT have accepted the need for logical thinking and active learning. The EFL teaching community needs to be aware of this: foreign or foreign-trained teachers are apt to encourage 'critical thinking', and many internationally developed EFL materials have an explicit 'critical' component (for example, chapter summary information in Pearson's *Password* series includes four headings: reading skills; target vocabulary; using critical thinking; and practicing writing), but the label has a negative connotation in Japan. The practical upshot of challenge one for teachers and curriculum designers trying to 'sell' CT goals to Japanese students or institutions is that they should consider using the language of logical thinking or (even better) active learning instead. Everyone can agree that improving comprehension and retention through active learning is good, while CT is for many a foreign niche skill.

Challenge two: traditional educational methods in Japan are not critical

According to the discussions in MEXT meetings, CT and active learning are particularly challenging in the Japanese educational context. Apparently, traditional educational methods in Japan have focused

on memorizing textbooks, reproducing correct answers, and one-way teaching (教員の教え込み, *kyōin no oshiekomi*) with little to no extended writing or discussion. Writing and class participation have typically had no direct impact on grades. An emphasis on CT therefore represents a significant change in classroom operation.

As early as 1986, MEXT experts were recommending a move to including written argument, debate, and report writing ('language arts') rather than only simple oral answers in Japanese class from primary and middle school; and discussion, debate, and writing of summaries and reports in English in high school. In 1998, updates to the 'Course of Study' curriculum guidelines were issued, emphasizing 'thinking' for the first time; these were updated in 2008 to emphasize thinking and 'judging independently'. Of course, an official recommendation does not necessarily result in real changes in classrooms, and discussion since suggests that change has been slow. External assessment did not change to reward these supposedly desirable skills, and many teachers did not themselves have the skills or time to teach or assess debates or written reports in Japanese, let alone English.

More recently, discussions from 2010 suggest that even if some high schools are sometimes having group discussions, they are not of a kind that helps critical thinking or problem-solving skills, and students are entering university 'extremely bad' at discussion. In a context where historically only the 'correct' view to be learned was presented, a teacher or textbook playing devil's advocate may be seen as endorsing an 'incorrect' view, and indeed, discussions from 2015 recorded on the MEXT site note that teachers don't know how to proceed in the face of a 'radical and risky' change to open discussion, and worry about the risk of complaints from parents or accusations of political bias (on Japanese culture's high level of risk-aversion see Hofstede 1983). Meetings in 2017 note that teachers desperately need training in leading discussions whereby high school students can discuss and appreciate different opinions and values.

By way of quantifying the changes over this period, a survey of nearly 1,300 Japanese high-school students in 2000 found that 90% never or seldom gave their opinion in class, and 60% never or seldom ask questions (Okada 2000). A repeat survey of nearly 1,500 students fifteen years later showed a substantial improvement to 62% and 40% (Okada 2015). While giving an opinion or asking a question is neither necessary nor sufficient for CT or active learning, the change suggests an increase in active involvement in classes. As a small international comparison though, the same survey administered to 64 Australian and 61 New Zealand high-school students in October 2018 (with thanks to two friends of the author) yielded answers between 10% and 20%. That is, between 80 and 90% of Australian students at least "sometimes" give their opinion and ask questions in class, compared to between 38 and 60% of Japanese students.

Indeed, in discussions at MEXT, the need to reform Japanese education is now characterized as ‘urgent’. Though presented in different ways by people from different backgrounds (subjects, school levels, and so on), a representative characterization is as an urgent need to shift educational and cultural paradigms (schools, curricula, teachers, and institutions) from ‘manufacturing things’ (for which imitation/memorization is a successful educational strategy) to an information economy which requires innovation and skills not acquired through imitation. Japan’s educational system has served it well enough historically, but classes focused on repetitive drills and evaluated entirely through multiple-choice questions are no longer fit for purpose.

The urgency and scale of these reforms are a major motivation for the 2020 overhaul of Japan’s hugely important ‘National Center Test for University Admissions’ into the ‘Common Test for University Admissions’ (大学入学共通テスト / *daigaku nyūgaku kyōtsū tesuto*) – an important enough change that in national media the new version is simply called the 新テスト (*shin tesuto* / new test; see for example National Center for University Entrance Examinations 2019). The new version is intended to assess powers of thinking, judgment, and expression rather than rote memorization, and thereby force changes to high school teaching methods (Timsit 2018, quoting a MEXT deputy director). In particular, where the old Center Test was entirely multiple-choice, the new test requires some short (from a word up to about four lines) written answers for the Japanese and mathematics sections; and where the old Center Test for English covered only reading and listening, the new test was intended to cover also speaking and writing. At the time of writing (November 2019), pushback from schools (ill-prepared to teach to these new English skills and concerned about marking) and other stakeholders has already forced a backdown on the English changes, delaying the inclusion of speaking and writing. Instead, certain private-sector standardized English tests which evaluate all four skills at once and have some track record and support infrastructure in Japan (including for example GTEC and TOEFL) were earmarked for use as external tests, but on the day that registration for these exams was due to start, MEXT backed down on this too, citing concerns over cost and access for students in remote locations, and uncertainty over fair comparison of different tests (see for example NHK 2019). While MEXT currently seems determined to go ahead with the short written answers in Japanese and mathematics, public pushback continues against these changes also. A representative piece from a popular magazine (Ishida 2019), suggests that 87% of stakeholders are worried about the new test format: convinced that accurate, confidential, and timely marking will be impossible (over half a million test-takers sit exams over two days each year); that graduate students are not qualified to act as graders; that students and teachers will be unable to grade performance on past or practice problems when the answer differs slightly from the model, making estimating scores impossible; that

unusual but excellent answers may be identified as incorrect, and so on.

As an aside, these concerns are not necessarily evidence of any unique lack in the Japanese system. For most of the history of the US SAT, only multiple-choice questions were used; essays were included only from 2005, and even now are purely optional. Analysis in 2005 suggested that essays were scored primarily by length (Winerip 2005), and since factual claims were irrelevant to grading, students were effectively encouraged to make up supporting evidence to create a particular rhetorical structure. A major redesign in 2016 switched to analysis of a supplied text, and introduced double-marking for all essays. In the UK, meanwhile, a study from The Office of Qualifications and Examinations Regulation examining marker variation at A-level (university entrance) shows that barely half of students receive the 'correct' (as determined by an experienced marker) exam grade in English Literature (Rhead *et al* 2016: 25). Marking for other 'subjective' subjects with essay-based assessment (History, Religious studies, Sociology, English, and so on) was only slightly more consistent, and in many cases assessment did not successfully spread out candidate marks, meaning that placement of grade boundaries (effectively, scaling) exacerbated the inconsistency. If countries with a far longer history of written assessment continue to have serious problems, reservations in Japan can be seen as reasonable caution instead of some kind of pedagogical Luddism. Internationally, these ongoing difficulties are motivating a significant increase in interest in automated scoring of student writing (see Burrows *et al* 2015). While Natural Language Processing and machine learning using large datasets of scored writing by English native speakers have yielded significant advances in recent years, automated scoring of high stakes assessment seems likely to remain highly contentious for some years (see for example Human Readers 2013).

To summarize, then, traditional educational methods in Japan are apparently not critical or active, but the situation has improved over the past 20 years, and there is an ongoing strong top-down push to better prioritize thinking, judgment, and expression over rote memorization. It is encouraging to note that pushback is centered around teachers being poorly prepared to teach CT, or the current centralized testing infrastructure being poorly prepared to assess it; pushback is not centered around CT being an inappropriate goal in Japan or an impossible goal for Japanese students. Neither area of pushback need be a problem for EFL teachers, who typically have some access to a more CT-friendly educational tradition and a different assessment situation (smaller groups, more time, and lower stakes). Indeed, CT and EFL seem to be a natural fit: using English helps enable a break from established classroom behaviors, and classes are smaller; added to which, languages classes work well when students are engaged and using target language meaningfully, even if to say something untrue, representing a CT-friendly focus on process over result.

Given this, the practical upshot of challenge two for language teachers in Japan seeking to increase the CT or active aspect of classes is essentially that students are likely to be unfamiliar with and perhaps nervous about target activities. Students may be used to teacher-centered classes aimed at teaching a single correct answer, and therefore unsure how to proceed with more open-ended, creative, or process-oriented activities. Fortunately, helping students who lack the confidence or ability to participate is the language teacher's stock-in-trade. Useful strategies to build student confidence include: modelling how 'mistakes' are a good way for the whole class to learn and improve; separating person and opinion or language ability; having students play a role; encouraging a supportive class culture; encouraging interactions between students; using ice-breaking activities; shuffling seating each class; providing clear models of target language; giving students time to prepare; designing activities to require repeated practice; and encouraging self- and peer assessment (see Bullsmith 2020 for details of these confidence-building classroom strategies).

Challenge three: little room for CT within the set curriculum goals or textbook

A more practical problem is that there may seem to be little room for CT when there is a detailed set curriculum which makes no mention of CT skills, or when classes are expected to closely follow a textbook with no explicit CT component. For teachers in such situations, it may be unclear how to encourage more CT, or indeed whether any class time should be spent on goals not on the curriculum.

The relationship noted above between CT and active learning is a useful starting point for answering both of these questions. That is, while CT may seem a difficult target, encouraging active engagement (evaluating, analyzing, synthesizing, summarizing, and so on) is more accessible. Likewise, while a teacher generally ought not depart from a detailed set curriculum to pursue different goals, particularly when the classes are part of some larger course or prerequisites for further courses, spending a little class time encouraging active engagement with the target material assists with understanding and retention.

As an example of how active engagement can be encouraged without departing from a set text and curriculum, effectively thereby also smuggling important CT skills into a course, I will report on several university classes taught using Oxford University Press's *Stretch 1* and *Stretch 2* (level A1-B1; Stemleski 2014a and 2014b) and some additional reading resources over two years. In the first year, I followed the teacher's guide and used most of the activities as given, taking notes on results, language produced, and so on. In the second year, running the same courses with new cohorts of students, I tried to encourage more active (read: critical) engagement. Some of the methods used were:

1. Soliciting or requiring questions as a summary/check of teaching points. Explain why questions are useful (for example, they help everyone, including the teacher). Elicit a range of questions about each teaching point ('What kind of questions can we ask about this?'), modelling examples as needed, and being liberal with praise. Typical questions extend the language given (opposites, etc.), place in context of prior knowledge (how is this different to x, could I say y instead, etc.), check use or collocation, use the language to express or check a fact, and so on. By requiring one or more questions about significant new language ('Hanako, can you ask a question for everyone?', 'Anyone—a question please!') students need to keep engaged so they can ask a question when needed. While many questions are display questions (the student knows the answer), the language use is still good practice, other students may learn from the exchange, and the stigma of asking a question when you genuinely want to check something is removed (volunteering a question starts being seen as a favor to classmates if it reduces the probability of them being called on to ask one). The teacher can add questions, modelling questioning and checking/deepening understanding; and when a supportive class culture has been established, student questions can be directed to the class with the teacher as backup/referee. The result: more actively engaged students who speak up more; a useful check on student understanding; asking questions changed from an embarrassment to a public good.

2. Recasting activities involving giving simple answers or opinions as micro-debates. Model some simple debate behavior and structures and explain that debating helps make sure you hear and use key language several times while thinking about the meaning, which is how we learn/remember/gain fluency. The goal can be to simply push your own position and criticize the other, or to reach agreement (and perhaps then report to another pair, in a pyramid/snow-ball activity). Start with assigned positions to lessen the shock of being contradicted and make criticism feel less personal, and to ensure that positions are different (and perhaps even funny, or apt to yield particular target language).

For example: an activity on adjectives ('Where would you most like to visit on vacation? Give three reasons. Share your ideas with a classmate.') yielded only five different destinations in the whole class (most chose Tokyo Disneyland, and 11 of 15 pairs agreed on destination). Each student read out their own notes while the other waited; little feedback, no excitement, limited range of language used. This was recast as 'The best place to visit on vacation is [Paris/Hawaii]. Give three reasons. Debate with your partner.'. The requirement to argue resulted in more attentive listening, interaction, negotiation: a noisier and more excited activity with a wider range of language produced. With the destinations defined, it's easy to swap and repeat, repeat with different destinations, or to choose destinations

to elicit particular language.

Or as another example: an activity on describing people ('Think about a friend or family member. What kind of person is he or she? Give examples. [Table column titles: Name/Personality/Examples] Write a paragraph about your friend or family member. Use your ideas in your chart above.') yielded only 'mother', 'father', and 'my high-school friend' as replies, with limited associated language. Again, each student read out their own notes while the other waited; little feedback, no excitement. Recast as 'Your new friend is [Donald Trump / Fukushi Sota]. What kind of person is he or she? Give examples. Write a paragraph about your new friend. Use your ideas in your chart above. Debate with your partner: who is the best friend?'. Again, students need to listen and understand, and the activity is noisier and more excited, and easy to tweak according to the interests and level of the students (new boss/love interest/neighbor) or with different famous/topical people to elicit particular language.

The result: more actively engaged students; more active listening; more language use and more target language use; practice disagreeing and being contradicted without risk of losing face; raised awareness that language practice needn't involve a single best/agreed answer.

3. Encouraging self and peer evaluation. The *Stretch* textbooks feature a basic form of self-evaluation, with each page ending with a space for students to check one of 'I can [use this key language/perform this task] ... Very well / Well / Not very well'. While such simple and immediate self-evaluation provoked little engagement in class (students seem to answer via the heuristic of substituting the simpler question of how they feel about the last micro-activity on the page for the more difficult and interesting question actually asked regarding the key language or task type in general), more demanding evaluation can prompt the kind of deep processing (recall and extended reflection) that Craik & Lockhart (1972) and active learning proponents argue improves learning outcomes.

Evaluation can be made more demanding by delaying it (what did you learn today; in the last class; in the last unit, and so on), by requiring it to be made with reference to a rubric (how was your partner's grammar; eye contact; relevance of content, and so on), or shared with others and thereby requiring some explanation or justification. In particular, evaluation with reference to a detailed rubric typically requires repeated reference to both rubric and the material to be evaluated, increasing exposure to target language and ideas.

For example: each unit in *Stretch 1* and *2* includes a short presentation on a prepared topic and using a particular presenting skill, in a format like: 'Stand up. Use your notes from Part C to tell a group about [the hobby/transportation you researched/etc.]. Remember to [use gestures/make eye contact/etc.].' In the first year, simply presenting in small groups as the textbook suggests resulted in

students spending most of their time either nervously preparing their own presentation while pretending to listen to others, or in a state of relief, casually nodding along with later presenters. Students were consistently unable to summarize the key points of even presentations that had just finished, even in later presentations where they must have known the teacher might ask, suggesting minimal attention. In the second year, self-and peer-evaluation was added, in the form of a mark sheet with a five-point rubric for each of content, clarity, eye-contact, and the presentation skill in that unit, on which each student evaluated themselves and received scores and comments from each person in their group. Sample praise and constructive criticism for each item was modelled. Evaluation sheets were collected and scores counted a small amount toward the course grade. The result: students paid closer attention, and were consistently able to summarize the key points of presentations in their group, even (with some prompting) when other presentations had been given since; students gained practice evaluating, comparing, giving and receiving praise and criticism. The teacher circulated around the class while presentations were being given, and managed to evaluate about 1 in 4 presentations through the year. Peer evaluations tended to be slightly higher than teacher evaluations, while self-evaluations tended to be slightly lower than teacher evaluations, but both were very highly correlated, suggesting that with care both can be utilized for assessment purposes, which could be particularly useful for larger classes (even with a modest class of 30 or so, serial teacher assessment of a short presentation by each student would likely take a whole class and become somewhat boring for students; with peer assessment, students practice in pairs before presenting in groups of four, followed by comments and debrief, and still finish in about a third of a class, no matter the class size).

4. Writing on the text / talking to the author / summarizing. This activity was not used with the *Stretch 1* and *2* classes, as the reading component is comparatively short and simple, but it proved a useful way to increase critical engagement with longer and more complex texts in an intermediate business English class, and serves as a useful contrast to the pair-work and groupwork ideas in #2 above. The idea is to write directly on the text, adding comments ('not in Japan!'), doubts ('really?'), structure (using arrows etc. to show logical structure and signposts), labels ('example', 'reason 2'), questions ('why?'), and so on. It's easy to try, can be set as part of reading homework, can be compared with a partner or group, works well as preparation for critical discussion of texts in class, and yielded some particularly interesting discussion when the set text was an advertisement. Some texts do lack an interesting logical structure or usefully disputable factual claims, and in that case, a more general fallback activity that works for most text types, whether reading or heard, is a simple summary. The need to produce a summary helps keep students engaged, requires constant evaluation (at

least of whether each point is a key point or a detail, with reference to the evolving understanding of the overall text), and requires the productive use of key language from the text.

Overall then, in response to challenge three, not having CT explicitly given as a curriculum goal or provided for in a set textbook need not be any barrier to encouraging it. Encouraging active learning—encouraging students to do something while learning, and think about what they’re doing (that is, to engage in deeper processing)—helps improve engagement, comprehension, and retention in language classes. This deeper processing, involving evaluation, judgment, comparison, and giving reasons, is also the groundwork of CT. It can be done with only small changes to classroom operation or textbook activities, for example by (1) eliciting student questions about teaching points, (2) recasting simple output activities as micro-debates, (3) encouraging self-and peer-evaluation, and (4) writing analysis and comments on texts or simply creating short summaries.

Student evaluations (on anonymized university-wide feedback forms) for the courses provided some interesting comments—the final questions ask for short written feedback on good points about the class, what could be improved, and other comments. In the first year, which included pair-work and presentations straight out of the text, the main theme of answers about good points was a friendly class atmosphere. In the second year, with a new cohort and involving pair-work recast as micro-debates and presentations subject to self-and peer evaluation, the main theme of answers about good points centered around interaction with classmates (‘difficult at first but so much fun; I made friends in class; I was nervous to start with but I ended up enjoying it; we didn’t discuss anything in high school, so it was great; I gained confidence through my classmates comments’, and so on). In terms of the student experience, at least, the more active classes were apparently better.

A small quiz at the end of each course asked students to name the eight important presentation skills practiced in class (score 0–8, half-points awarded). The second cohort mean score was 1.76 higher than the first cohort mean score (a statistically highly significant difference, *t*-Test $P < .001$). With very similar class time spent on presentations in each cohort, this difference is credibly due to the deeper processing and repeated exposure while carrying out four evaluations using a rubric including the skill.

Conclusions

Firstly, CT is an unpopular label in Japan, and while senior educators and policy-makers generally want it better promoted, educators trying to promote CT within a school may wish to use the labels ‘logical thinking’ or (even better) ‘active learning’ instead.

Secondly, students in Japan are likely to be unfamiliar with and perhaps nervous about CT or

'active learning' activities. Various strategies and activities designed to increase student confidence were discussed.

Thirdly, while there may seem to be little room for CT within set curriculum goals, spending a little time adding a CT slant to activities need not be seen as taking time away from existing goals. Rather, encouraging learning through active tasks such as evaluating and summarizing should improve student engagement and learning outcomes toward the set curriculum goals, and is therefore pedagogically desirable independent of any desire to promote CT.

References

- Bullsmith, C. (2020). Helping Japanese University Students Gain Confidence. 跡見学園女子大学 [Atomi University FD Journal], 18.
- Burrows, S., Gurevych, I., & Stein, B. (2015). The Eras and Trends of Automatic Short Answer Grading. *International Journal of Artificial Intelligence in Education*, 25(1): 60–117.
- Craik, FIM. & Lockhart, RS. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. DOI:10.1016/s0022-5371(72)80001-x
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. California Academic Press. Retrieved from <https://philpapers.org/rec/FACCTA>
- Hofstede, G. (1983). Culture's Consequences: International Differences in Work-Related Values. *Administrative Science Quarterly*. Johnson Graduate School of Management, Cornell University. 28 (4): 625–629. DOI:10.2307/2393017
- Human Readers. (2013). Retrieved from <http://humanreaders.org/>
- Ishida, K. (2019). 大学共通テスト記述式「大学生が採点かも」「自己採点できない」に悲鳴 [Common Test for University Admissions written answers: outcry over 'university students may work as graders', 'students can't work out their own marks']. *Aera*, 28 September 2019. Retrieved from <https://dot.asahi.com/aera/>
- Ministry of Education, Culture, Sports, Science and Technology. 文部科学省ホームページ [MEXT Homepage]. Retrieved from <http://www.mext.go.jp/>
- National Center for University Entrance Examinations. 大学入学共通テスト [Common Test for University Admissions]. Retrieved from https://www.dnc.ac.jp/daigakunyugakukibousyagakuryokuhyoka_test/
- NHK (2019). 英語民間試験来年4月からの実施を見送りへ [April Implementation for Private English Tests Delayed]. NHK News Web, 1 November 2019. Retrieved from <https://www3.nhk.or.jp/news/>
- Okada, R. (2000). Critical thinking in the Japanese classroom. *Proceedings of the Tokai University Junior Colleges*, 33, 1–9.
- Okada, R. (2015). Thinking in the Japanese Classroom. *Journal of Modern Education Review*, 5(11), 1054–1060. [https://doi.org/10.15341/jmer\(2155-7993\)/11.05.2015/004](https://doi.org/10.15341/jmer(2155-7993)/11.05.2015/004)
- Rhead, S., Black, B., & Pinot de Moira, A. (2016). Marking consistency metrics. Ofqual. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/681625/Marking_

consistency_metrics_-_November_2016.pdf

Stempleski, S. (2014a). *Stretch: Level 1: Student's Book with Online Practice*. Oxford University Press.

Stempleski, S. (2014b). *Stretch: Level 2: Student's Book with Online Practice*. Oxford University Press.

Timsit, A. (2018). Overhauling Japan's High-Stakes University-Admission System. *The Atlantic*, 13 January 2018.

Retrieved from <https://www.theatlantic.com/>

Twardy, C. (2004). Argument Maps Improve Critical Thinking. *Teaching Philosophy*, 27(2), 95-116. <https://doi.org/10.5840/teachphil200427213>

Winerip, M. (2005). SAT Essay Test Rewards Length and Ignores Errors. *The New York Times*, 4 May 2005. Retrieved from <https://www.nytimes.com/>