

Critical thinking efficacy and transfer skills defend against ‘fake news’ at an international school in Finland

Journal of Research in
International Education
2019, Vol. 18(1) 23–41
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1475240919830003
journals.sagepub.com/home/jri



Shane Horn

Amman Bacculaureate School, Jordan

Koen Veermans

University of Turku, Finland

Abstract

In this study, tasks measuring digital media literacy developed by Stanford University were administered at a school in Finland to consider the efficacy and transfer of critical thinking (CT) skills of a ‘pre-IB’ cohort preparing to enter the two year International Bacculaureate Diploma Programme (IBDP) and a graduating ‘IB2’ cohort. While the IB2 cohort outperformed the pre-IB cohort, both outperformed Stanford’s U.S. cohorts to a statistically significant degree. Utilising a framework of curricular approaches to facilitating CT skills development as a variable of interest for causal-comparison, it was determined that the Finnish curricula and the IBDP explicitly facilitate CT skills as a separate course while embedding CT into subject coursework, whereas the curriculum in the U.S. implicitly embeds CT into subject coursework only. Implications for improving facilitation of CT in curricula design, professionalising CT across the field, and the benefits of replicating existing studies in differing socio-educational environments are discussed.

Keywords

International Bacculaureate, Finnish education, critical thinking, media literacy, fake news

Introduction

A recent article in *Foreign Policy* considered the ways in which false news stories designed to aggravate existing social problems spread on social media, with the author contending that the strength of Finland’s education system has led to ‘widespread critical thinking skills’ (Standish, 2017: para. 10) which have effectively equipped the population to defend itself

Corresponding author:

Koen Veermans, University of Turku, Finland.

Email: koevee@utu.fi

against disinformation. Similar sentiments were expressed by Finland's Director of Communications, who claimed that Finnish citizens have been notably successful in their ability to discern credible information online: 'We have trolls. We have fake news. But the Finns do not buy false news', he said, referring by way of explanation to the Finnish education system's focus on developing critical thinking (CT) skills in their learners (Bee et al., Allen and Pennolino, 2017). These causal claims appear to assume a correlation with Finland's reputation for excellence in public education.

While the research described in this article does not deal directly with claims of intergovernmental propaganda and campaigns of disinformation, it does consider the thinking skills involved in evaluating the reliability of evidence found in online contexts and the role of educational curricula in helping to develop those skill sets. The study is broadly concerned with CT efficacy, or the extent to which the facilitation of CT skills may be producing its intended result. More specifically, it considers the ways in which CT skills explicitly developed through educational curricula at an International Baccalaureate (IB) Diploma Programme (DP) school in Finland may transfer to external contexts of daily online interactions.

Literature Review

Increased CT skills development is 'widely cited by national education groups, teacher unions, higher education organizations, and workforce development groups as an imperative for today's students' (Silva, 2009: p. 630) and is reflective of a global push across various national curricula toward this end. This 'push' is most visibly found in Australia (Australian Curriculum, Assessment and Reporting Authority, n.d.), Canada (Fillion and Martelli, 2017; Ontario Ministry of Education, 2017: p. 8), England (Glevey, 2008), Singapore (Leen et al., 2014) and some states in the U.S. (Ennis, 2018: p.165; Gewertz, 2008; Silva, 2009: p. 630) as an explicit component of the educational objectives. Yet little work appears to have been conducted in measuring the outcomes of such efforts (Tiruneh et al., 2014: pp. 1-2) to consider the extent to which they are fulfilling their intended purpose. The research described here aims to contribute toward filling that gap.

Defining critical thinking

There emerges a pattern within the existing academic literature wherein the concepts, qualities, or processes necessary to establish a universal working definition for CT are considered difficult to determine (Ennis, 1991; Larsson, 2017; McPeck, 1990; Orszag, 2015; Paul, 1995; Saiz and Nieto, 2010; Tiruneh et al., 2014), a problem which naturally results in CT being defined variously. Despite this common concession and problematised viewpoint, there emerge further patterns toward shared defining parameters, which nearly every study then proceeds to list with some semblance of the historicity involved in determining CT's contemporarily accepted meanings. The current research necessitates establishing that a working definition and conceptual understanding of CT fit within the qualities being measured by the context and purpose of the study. To this end, in the methods and materials section of this article the task instruments will be explored and correlated to common accepted meanings of CT.

Critical thinking efficacy. A central driving question to this research asks: 'How do we know the explicit facilitation of CT skill sets is developing the desired outcomes in learners?' The exploration applies the criteria and approach of CT to the teaching and learning of CT, promoting fundamental questions about the ways in which the skill sets ideally developed through CT instruction translate into real world CT outcomes. From a working understanding of efficacy to mean 'the

ability to produce a desired or intended result' (Simpson and Weiner, 2018), the term 'critical thinking efficacy' seems to best encapsulate the overriding concept under investigation. While the meta-analyses of various intervention studies in CT and recent scholarship on media literacy reveal that research into CT efficacy has begun, searches for the phrase in both EBSCO and the International Education Research Database revealed no relevant returns with the words in that order with that specified or intended meaning.

This may be indicative of the recentness with which CT efficacy has become an active concept in academia. For a working definition, 'critical thinking efficacy' denotes consideration of the extent to which the facilitation of CT skills produces the desired or intended results of its efforts. The definition is broader than the scope of this study, as it does not limit the term to the explicit facilitation of CT by external actors and can include 'self-taught' CT skills development. The former remains the more specified subject of CT efficacy under present investigation.

Critical thinking transfer. The central role of transfer with CT skills facilitation appears throughout its varying definitions and is particularly well highlighted in Tiruneh et al. (2014), where they state that:

CT instruction is mainly based on the assumption that there are identifiable and definable thinking skills which are domain-independent and can be taught to students to recognize and apply them appropriately in daily life situations and future careers. The goal of CT instruction is, therefore, to help students acquire and transfer those domain-independent thinking skills to solve problems faced in everyday life. (p. 3)

The precedents for this determination can be found in Glaser (1984), Ennis (1989), Halpern (1988)—whose research into teaching CT for transfer across domains provided focus on 'training in the structural aspects of problems and arguments to promote transcontextual transfer of critical-thinking skills' (p. 449), Nickerson (1988), and Perkins and Salomon's 'Teaching for transfer' (1988). These scholars contend that the issue of transfer must be addressed whether CT is taught using a general or discipline-embedded approach.

Nickerson (1988) noted that teaching CT separate from content runs the risk that learners gain some comprehension of CT principles but may 'fail to connect that knowledge to the many situations in life in which it could be useful', and that a risk of teaching the same aspect of thinking only within the context of a standard course is that a student will 'fail to abstract from the situation what is really context independent and again will not transfer what has been learned to other contexts' (p. 34). This study shares the assumption that CT development revolves around the issue of transfer, and aligns with the leading research by focusing on the extent to which skills facilitated in CT-oriented curricula effectively transfer to evaluating information and media literacy tasks based on situations faced in daily interactions with social media and online news.

Information and media literacy

The materials utilised as measurement tools for this study consider the development of CT skills which fall under the canopy of 'information literacy' generally, 'media literacy' more specifically, and even more specifically within digital contexts representative of young people's daily interaction with social media and online news.

Indicative of the timeliness of the topic to the advent of 'fake news' permeating current news cycles, during the course of this study a publication titled 'The promises, challenges, and futures of media literacy' by the Data & Society Research Institute offers an evaluation of media literacy efforts to date with contextualisation to the current landscape. The authors state that 'In general,

there is a lack of comprehensive evaluation data of media literacy efforts' (Bulger and Davidson, 2018: p. 3-4), beginning their exploration by stating that media literacy is conceived 'as a process or set of skills based on critical thinking' (p. 3); they directly tie their working definition of media literacy to a 'skill set that promotes critical engagement with messages produced by the media' (p. 4). For a general working definition, media literacy is the 'active inquiry and critical thinking about the messages we receive and create' (Hobbs and Jensen, 2009), further emphasising the connection to CT skills development.

The authors quote researchers Wineburg and McGrew who developed the task instruments utilised in this study, wherein they offer commentary on the dangers inherent to the internet opening 'the floodgates to misinformation, fake news, and rank propaganda masquerading as dispassionate analysis' (cited in Bulger and Davison, 2018: p. 5). The report concludes with a series of open questions, some of which overlap with the research questions and implications of the present study. Questions such as 'Can media literacy even be successful in preparing citizens to deal with "fake news" and information?' and 'Are traditional media literacy practices (e.g., verification and fact-checking) impractical in everyday media consumption?' (p. 21) show similar inquiry into CT efficacy.

Ennis' curricular approaches for teaching critical thinking

A categorisation scheme for curricular approaches to teaching CT developed by Ennis (1989), and explored, defined, and utilised more recently by Tiruneh et al. (2014) in their meta-analysis of CT intervention studies at the higher education level, serves as a pragmatic framework. In this framework, the *general* approach is defined as when CT is explicitly taught in a separate course from other subjects. The driving assumption for this approach is that the skills developed will naturally transfer for use in other disciplines. The *infusion* approach integrates CT instruction with standard subject matter instruction while making general principles of CT explicit to learners. In this approach, students are encouraged to 'acquire and explicitly practice CT skills through deep and well-structured subject matter instruction' (p. 2). The *immersion* approach also integrates CT skills within standard subject matter instruction, but general CT principles and procedures are not made explicit. The assumption with this approach is that learners will naturally acquire the thinking skills from engaging in CT-oriented subject instruction without studying it separately. The *mixed* approach consists of a combination of the general approach with either infusion or immersion. In a mixed approach 'there is a separate thread or course aimed at teaching general principles of CT, but students are also involved in subject-specific CT instruction where either the objectives of CT are explicit or implicit' (p. 3).

A natural question to ask is which approaches are more measurably effective than others when compared; an inquiry which does not yet yield a concrete conclusion. The more answerable quantification is determining which approaches are most commonly utilised. Systematic reviews of CT instructional approaches (Abrami et al., 2008; Behar-Horenstein and Niu, 2011; Tiruneh et al., 2014) reveal that over three-fourths of intervention studies utilise either immersion or infusion. Despite their popularity, Tiruneh et al. (2014) found students' abilities to transfer acquired CT skills from subject-matter instruction to new tasks equally ineffective via either of these approaches. While the general and mixed approaches were found to more effectively enhance CT skills development, the researchers caution that the smaller number of studies on these approaches may limit their generalisability.

Abrami et al. (2008) reached a different conclusion by finding the immersion approach less effective than the infusion approach, and indeed the variable upon which there is more agreement is whether the CT skills being developed are made explicit (Ennis, 1989; Mayer, 1992; McPeck,

1990; Nickerson, 1988; Resnick, 1987), such as would be found in the general, infusion, and mixed approaches. The available evidence suggests that ‘direct teaching strategies, which are based on explicit and detailed explanation of CT principles, are more effective than the implicit teaching strategies’ (Tiruneh et al., 2014: p. 8). The evidence on the overall effectiveness of implicitly embedding CT skills such as asking higher-order questions, concept mapping, and facilitation of small group discussion within subject matter instruction without any explicit instruction in CT development remains inconsistent and inconclusive (pp. 5-6).

Curricula under consideration

The student participants in this study consist of a ‘pre-IB’ cohort (n=42) preparing to enter the two year pre-university International Baccalaureate (IB) Diploma Programme (DP) and an ‘IB2’ cohort (n=25) preparing to graduate from the programme. The majority of these participants studied within the Finnish education system prior to entering this programme (see Demographic Data section below) and the results from the pre-IB and IB2 students were compared with those from public schools in California. Each of these three guiding curricula are therefore analysed here with regard to Ennis’ (1989) categorisation for CT instruction.

International Baccalaureate Diploma Programme. The IBDP’s overall design is mixed by combining the general with infusion approach. The compulsory Theory of Knowledge (TOK) course is specifically designed to facilitate development of CT skills which are taught explicitly as a separate subject from the other academic disciplines (general). The subject guides for each course within the IBDP curriculum include connections to TOK which are made to be taught and assessed explicitly by subject area teachers (infusion).

While this determination applies to the curricular approach, it is possible that a randomly selected IBDP subject area classroom anywhere in the world could be found to be following the immersion approach in practice; for example, if an IBDP subject teacher does not explicitly cover TOK in the course instruction. This was likely to be more common under previous course guides, but recent revisions to the IBDP have integrated TOK into the content and assessment structures to the extent that the current curriculum treats all subject teachers as also being teachers of TOK, thereby aligning it with the mixed infusion approach in terms of its curricular design, structure, and intended implementation.

Finnish National Core Curriculum. The design of the Finnish National Core Curriculum (NCC) is also mixed, but by way of combining the immersion approach with the general. The compulsory courses in worldview and ethics throughout the NCC and the philosophy course at the upper secondary level are taught as explicit CT courses separate from the other academic disciplines (general). Meanwhile, the NCC makes clear that CT skills are at least implicit across certain subject areas: the specific term ‘critical thinking’ is found twenty times embedded in over ten different courses in the NCC for Basic Education, and fourteen times embedded in over ten different courses in the NCC for General Upper Secondary Education (immersion). While Finland’s curricular structure follows mixed immersion, there remains the possibility of meeting criteria for mixed infusion at the classroom level dependent upon the individual teacher, who in Finland retains autonomy over instructional approaches to learning.

California Common Core State Standards. The term ‘critical thinking’ does not appear in the Common Core State Standards (CCSS), nor is there a separate course within the curriculum which aims to facilitate CT skill sets in isolation of the subject areas as per the general approach. However,

there is evidence of CT skills development being implicitly embedded into the curricular frameworks as per the immersion approach. In the curricular frameworks, the term ‘critical thinking’ is found 30 times in history/social science and 20 times in English language arts/English language development. The latter has a chapter dedicated to ‘learning in the 21st century’ that includes a section titled ‘critical thinking skills’ and further sections dedicated to fostering global awareness, digital citizenship, and understanding multimedia text (CADOE, 2015: p. 937). It is unclear how these skills are integrated into the curriculum while absent from the CCSS. As with Finland, there remains the possibility that subject teachers could cover CT skills explicitly in their coursework to follow an infusion approach for those individual classrooms.

Materials and Methods

An inspiration for the overall design and approach of this study came from the meta-analysis, ‘Facts are more important than novelty: Replication in the education sciences’ (Makel and Plucker, 2014), which analysed the top 100 peer-reviewed education journals and found that only .13% of articles were replications. While most revealed results supportive of the original studies, they were less likely to replicate successfully when authorship differed between the original and replicating articles (2014).

While this study is not a direct third-party replication as per Makel and Plucker’s suggestion, since the materials were administered on a differing demographic, it does directly replicate materials in similar fashion to the original study. The benefit of replicating the same task instruments in differing socio-educational contexts and curricula is that it helps researchers determine the strength of generalisability of the implications from the original study. That is, if a replicated study reveals similar results, the generalisability of the original study is strengthened, whereas if a replicated study differs in its results then there exists a threat to overall generalisability. The extraneous variables introduced in this replication are: 1) the age and grade level of one class graduating from the IBDP in Finland versus another class preparing to enter the IBDP, and 2) the socio-educational curricula amongst the U.S. cohorts and the IBDP school in Finland’s two cohorts combined. As such, these remain the variables of interest for further consideration of the results.

Task assessments used to measure the extent to which students exhibit media literacy skills in online contexts were requested from the Stanford History Education Group (SHEG), which provided access to them for replication. The tasks and results were made available to the public at the time of the SHEG study’s publication (McGrew et al., 2018); however, when the measurement in Finland was conducted the SHEG had only publicly released one task from the high school level with limited information on their results in an executive summary (Wineburg, et al., 2016). This was expressly published in some media, but the remaining four tasks at the high school level were kept confidential. In addition to providing access to these tasks and evaluation rubrics, the SHEG provided additional information on their results under the condition that the data not be publicly shared until publication of their January 2018 journal article (Breakstone, 2017).

The SHEG study received widespread media attention due to its completion coinciding with the advent of the ‘fake news’ phenomenon (Domonoske, 2016; Schulten and Brown, 2017; Shellenbarger, 2016; Wineburg and McGrew, 2016). Stanford’s study was descriptive in nature and covered much by way of breadth, whereas the present study reaches more depth by measuring outcomes between groups within curricula designed to explicitly develop CT skills. Further, the present study offers comparative value by contrasting the performance outcomes collected by the SHEG of high school students in the U.S against the two high school cohorts measured in Finland.

In their systematic review of intervention studies on CT instruction, Tiruneh et al. (2014) conclude that evaluation of the effectiveness of CT instruction could be influenced by the type of

Table 1. High school assessment task descriptions with SHEG sample size.

Task	Topic	Description	SHEG Sample Size
1	Argument analysis	Students determine which of two arguments responding to an online article about a social media application is stronger.	172
2	News on Facebook	Students explain which of two similar news posts on social media is a better source.	173
3	Facebook argument	Students explain which poster in a social media conversation provides stronger evidence in a debate about gun laws.	175
4	Evaluating evidence	Students evaluate the strength of evidence in a photograph posted on an image-sharing site.	170
5	Comparing articles	Students explain which of two articles on a major news website is a more reliable source about climate change.	176

measures employed in a study. For example, two infusion approach studies by Anderson et al. (2001) and Bensley and Haynes (1995) reported differing outcomes when they utilised the same teaching strategies and research design but differed in CT measurements. The authors noted that some variations on CT outcomes could be explained by the multiple-choice format of the standardised CT measure being utilised. In a study by Plath et al. (1999) in which two CT measures were utilised together, significant CT improvement was revealed on the measure that required students to respond to open-ended items rather than in a multiple-choice format (Tiruneh et al., 2014: p. 8). In addition to employing a standardised assessment with open-ended items, the present study reduces CT measurement type as a confounding variable by replicating assessment tasks from a prior study.

Stanford History Education Group task development

The assessment tasks were developed by the SHEG over three phases of an 18-month period which covered 12 states and led to the collection of 7,804 responses at the middle school, high school, and college levels. Sites for field-testing included under-resourced, inner-city schools as well as better-resourced schools in the suburbs. Five assessments were finalised for each level. The SHEG addresses the issue of using paper-and-pencil tasks to measure digital literacies with an OECD study (2015) establishing that important abilities for judging online sources can be effectively measured offline (Wineburg et al., 2016: p. 6).

The final high school level assessment tasks were administered to 348 students in participant groups of between 170 and 176 across three districts in California. The districts had diverse populations with a free and reduced student lunch rate of 36%, 55%, and 68%. Students were given 30 minutes to complete packets of three tasks, which were randomly divided so that half of the students in each class completed one packet of three tasks while the other half completed a packet of three different tasks (McGrew et al., 2018: p. 8). Assessment rubrics were created by the SHEG with categories determining performance levels of beginning, emerging, or mastery. Descriptors were included at each level for an assessor to identify, evaluate, and categorise student responses. Sample responses with a discussion on the elements which placed the response at the corresponding level were provided for each performance outcome level.

Table 1 shows the final task topics with brief descriptions and the corresponding SHEG sample size (Breakstone, 2018). The task instruments and rubrics have been made publicly available for classroom use since being utilised in this study and can be accessed with the creation of a free account through the SHEG website (<https://sheg.stanford.edu/civic-online-reasoning>).

The SHEG use the term ‘civic online reasoning’ to differentiate the set of practices developed through their assessment tasks from broader understandings of media literacy, which might include separate competencies from learning how to type to advanced programming and coding. To the SHEG, civic online reasoning is ‘a more narrowly focused term to describe how to evaluate and use online information to make decisions about social and political matters than the larger field of media literacy’ (McGrew et al., 2018: p. 5) such as those found in descriptions from the National Association for Media Literacy Education (2007) and the National Council for the Social Studies (2016). For the purposes of this paper, given the established understandings of media literacy to include development in CT skill sets, the terms civic online reasoning and information, media or digital literacy are transposable.

The SHEG tasks are designed to assess students’ abilities to evaluate the reliability of sources of information and consider the role of evidence in arguments within online contexts, with the driving categorical questions at the high school level of ‘Who is behind the information?’ and ‘What is the evidence?’ well within existing conceptual and defining properties of CT. The assessment goals of these tasks align with understandings of CT from Dewey’s (1910) first contemporary conception as providing evidential and reasoned judgments, to Glaser (1941) interpreting and evaluating arguments, to the expert consensus definition generated for the Delphi Report (Facione, 1990) and nearly every preceding definition and conceptual understanding of CT to the present day.

Validity

In the prototyping phase, the SHEG researchers began by administering 56 tasks in a product design method which sought user testing for revision and improvement. In the validation phase, extensive piloting was undertaken and qualitative data through ‘think aloud’ interviews were collected by hundreds of participants to establish cognitive validity, defined as ‘the relationship between what an assessment seeks to measure and what it actually does’ (Wineburg et al., 2016: p. 5). During the final phase of field testing, thousands of responses were collected along with teacher consultations until 15 assessments were finalised, with the SHEG concluding: ‘Together with the findings from the cognitive validity interviews, we are confident that our assessments reflect key competencies that students should possess’ (p. 5).

Reliability

A secondary rater familiar with the testing materials was trained on the assessment rubrics, and evaluated a 20% sample distributed evenly across the five tasks from both cohorts to compare against a primary rater who evaluated the full allocation of tasks. Inter-rater agreement was $\alpha = 0.86$ for the pre-IB cohort, $\alpha = 0.90$ for the IB2 cohort, and $\alpha = 0.89$ for the pre-IB and IB2 combined, where $\alpha > 0.80$ indicates a strength of agreement with a high probability of returning similar results from other external raters (Krippendorff, 2018).

Results

Administration of tasks

Tasks were administered during two different testing sessions over the winter of 2017 at a school in Finland utilised by the affiliated university for teacher training and research. The first cohort measured, the IB2 (n=25), were students in their second year and nearing completion of their IB DP studies for high school graduation. The second cohort measured, the pre-IB (n=42), were in an IB

preparation programme. Both administrations took less than an hour with time allocated for reading a standardised introduction, task completion, reading and explaining a consent release form, discussing the tasks, and then answering any questions about the nature of the study. All 67 student participants provided consent for inclusion of their data. During these post-administration discussions, it was established that none of the participants had previously seen the specific tasks or read about the SHEG study.

Demographic data

The cohorts were nearly two thirds female in the pre-IB and nearly three fourths female in the IB2. 62% of the pre-IB cohort indicated Finland as their country of origin, or 74% when combining those who responded with both Finland and an additional country; the remaining 26% were spread evenly amongst individual or pairs of students from nine different countries. 76% of the IB2 cohort responded with Finland as their country of origin, with the remaining 24% spread evenly amongst individuals from six different countries. 52% of the pre-IB group listed Finnish as their first language, or 76% when combining those who responded with Finnish and another language; the remaining 24% were spread amongst individual or pairs of students from ten different first language backgrounds. 52% of the IB2 group listed Finnish as their first language, or 76% when combining those who responded with Finnish and another language; the remaining 24% were spread amongst individual or pairs of students from six different first language backgrounds.

62% of the pre-IB cohort indicated they had their entire schooling within the Finnish education system, with 76% having experienced five or more years within the NCC. Fewer than 20% had between one and four years' experience, with only one participant having no previous exposure at all. 72% of the IB2 cohort indicated that their entire schooling experience was within the NCC, with 92% having experienced five or more years; the remaining 8% were two non-responses. The IB offers two other programmes for younger age ranges: the Primary Years Programme (ages 4-11) and Middle Years Programme (ages 11-16). When asked about prior exposure to studying in an IB programme, over 95% of the pre-IB group had no such exposure, with one participant indicating four years' previous experience and another indicating the student's entire schooling to date had been within IB curricula. 64% of the IB2 group indicated two years' experience in the IB, with an additional 24% indicating three years, for which they may have been including their pre-IB year, which would indicate 86% with no previous exposure. Three participants, or 12% of the IB2 cohort, indicated four years of previous IB experience.

Post-survey data

In the post-survey, 78% of the pre-IB respondents indicated that they had knowledge of the 'blue checkmark' (measured in Task 2) utilised by social media networks to verify that companies, celebrities, news outlets and journalists are who they say they are, while 90% of the IB2 respondents indicated such awareness. 88% of the pre-IB group and 71% of the IB2 group indicated that they use social media several times a day, with 93% and 88% respectively indicating their use of social media at least once daily. Only one respondent across the two cohorts indicated using social media once a week or less, and none of the respondents indicated that they do not use social media at all. Respondents typically listed several sites such as Instagram, Twitter, Facebook, Reddit, and YouTube amongst the social media networks they most frequent.

Fewer students from both cohorts indicated that they read or watch news media online, with almost half the pre-IB (47%) group and just over three-fifths of the IB2 group (67%) reporting their frequency of use as at least once a day, and just over 40% of the pre-IB group and 29% of the IB2

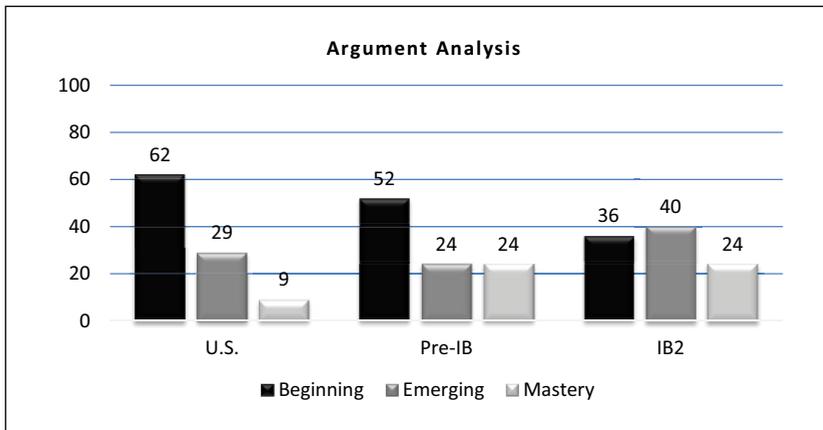


Figure 1. U.S., pre-IB, and IB2 results on Task 1.

group who indicated consuming news media online a few times a week. Only 6% of both cohorts indicated their frequency of online news consumption as less than once per week, with two students from the pre-IB and none from the IB2 group indicating that they never consume news media online. Most students listed multiple sources for their news such as CNN, BBC, *The New York Times*, and prominent Finnish sources such as the *Ilta-sanomat* and *Iltalehti* at the national level and the *Helsinginsanomat* and *Turunsanomat* more locally, as well as Yle the public broadcasting service. Many participants indicated familiarity with topics such as knowing Donald Trump or *The New York Times*, while none indicated prior awareness of the tasks.

Results on tasks

While none of the comparisons for individual tasks revealed significant differences under non-parametric Mann-Whitney U testing, comparisons between the pre-IB and IB2 groups when combining the tasks revealed a significant difference with $p = .045$ (or, if Task 2 is removed on account of it largely measuring knowledge over thinking skills, $p = .022$ for the remaining four tasks combined). The lack of significant outcomes for the individual tasks between the two cohorts could be explained by the sample sizes in combination with the fact that there are only three possible outcome values.

Results per task from the SHEG study, which were expressed only in percentages, are shown alongside the pre-IB and IB2 results. Although the raw data from the SHEG are not available, an indication of the differences between groups can be obtained from χ^2 comparisons which reveal pattern differences. These differences can then be qualitatively interpreted based on the pattern of the distributions in the different groups. Since no data were available on individual student responses from the SHEG, composite score comparisons were not conducted. Distributions of the U.S. cohorts were created based on percentages and reported sample sizes for the different tasks, where the number of task participants is multiplied by the percentage results per task and then rounded to achieve a real number to account for the differential in sample sizes between groups.

On Task 1 (Figure 1), the pre-IB and IB2 performed at lower beginning levels compared to the U.S. results, with the pre-IB performing lower and the IB2 performing higher at emerging, and

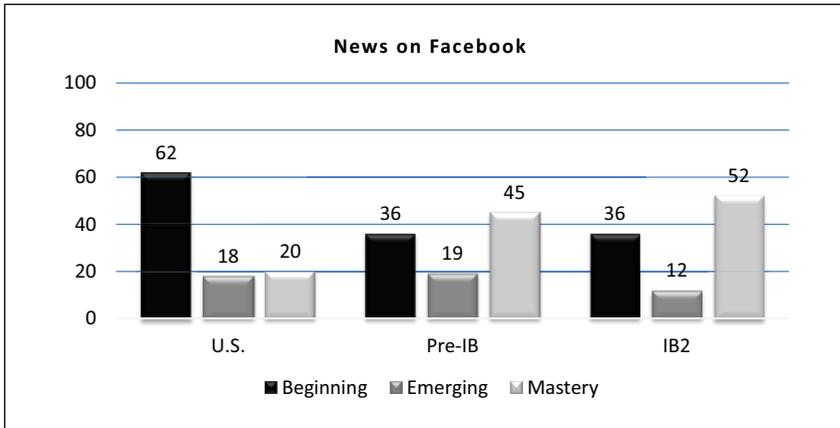


Figure 2. U.S., pre-IB, and IB2 results on Task 2.

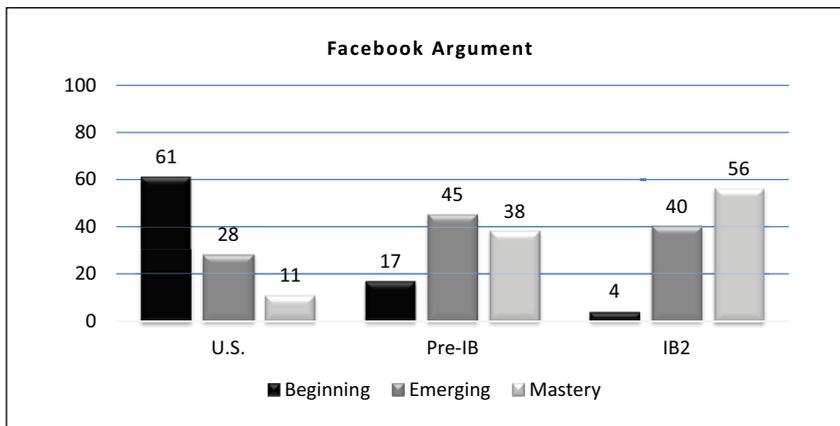


Figure 3. U.S., pre-IB, and IB2 results on Task 3.

both performing equally higher at mastery. The differences between the pre-IB and the U.S. ($\chi^2 = 7.45, p = .02$) and the IB2 and the U.S. ($\chi^2 = 8.18, p = .02$) are statistically significant.

The pre-IB and IB2 performed at equally lower beginning levels than the U.S. on Task 2 (Figure 2), with the pre-IB performing slightly higher and the IB2 performing lower at emerging; both performed higher at mastery. The differences between the pre-IB and the U.S. ($\chi^2 = 12.50, p = .00$) and the IB2 and the U.S. ($\chi^2 = 12.17, p = .00$) are statistically significant.

On Task 3 (Figure 3) the pre-IB and IB2 cohorts performed at lower beginning levels and higher emerging and higher mastery levels than the U.S. group. The differences between the pre-IB and the U.S. ($\chi^2 = 31.55, p = .00$) and the IB2 and the U.S. ($\chi^2 = 41.31, p = .00$) are statistically significant.

The pre-IB and IB2 performed on Task 4 (Figure 4) at lower beginning and higher emerging and mastery levels than the U.S. The differences between the pre-IB and the U.S. ($\chi^2 = 29.48, p = .00$) and the IB2 and the U.S. ($\chi^2 = 41.76, p = .00$) are statistically significant.

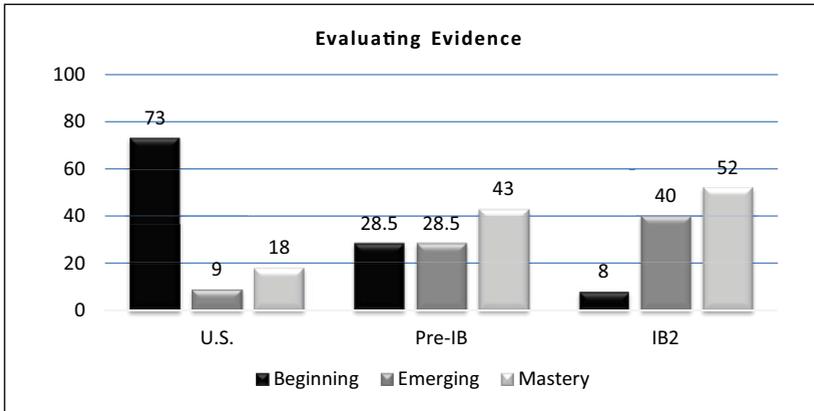


Figure 4. U.S., pre-IB, and IB2 results on Task 4.

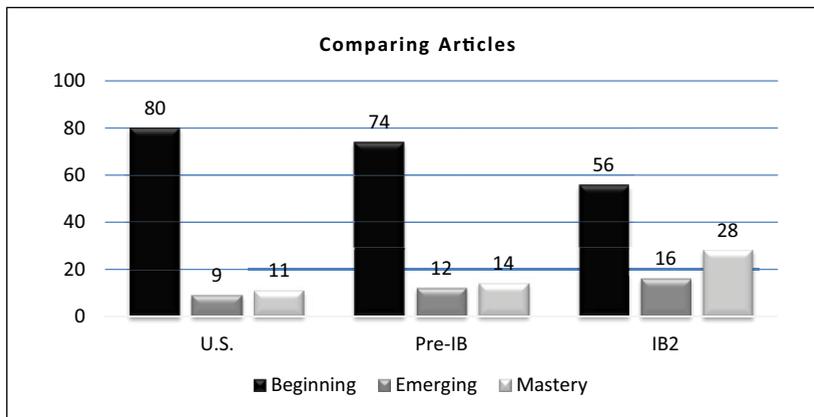


Figure 5. U.S., pre-IB, and IB2 results on Task 5.

On Task 5 the pre-IB group performed slightly lower and the IB2 performed lower at the beginning level than the U.S., and both performed higher at emerging and mastery levels (Figure 5). Although the difference between the pre-IB and the U.S. cohorts is not significant ($\chi^2 = .81, p = .67$), the difference between the IB2 and the U.S. is statistically significant ($\chi^2 = 7.71, p = .02$).

On Tasks 1, 3, and 4 (designed to ask ‘What is the evidence?’), the IB2 results revealed 8% of the cohort who performed at the beginning level on two of these tasks and none who performed at the beginning level on all three, whereas the pre-IB had over 20% of the cohort performing at the beginning level on at least two of these three tasks, and 7% at the beginning level on all three. Conversely, whereas 60% of students in the IB2 cohort did not perform at the beginning level on any of these three tasks, this percentage was half that at 30% for the pre-IB. While over 20% of the pre-IB cohort did not achieve mastery level on any of the five tasks, all students in the IB2 cohort achieved mastery on at least one task. Only two of the 67 participants achieved beginning levels on all tasks, both of whom were in the pre-IB cohort.

Discussion

The conclusions of the SHEG in their study offer a strong warning that CT efficacy is largely not occurring amongst the students measured in the U.S. ‘Overall’, the authors state, ‘young people’s ability to reason about information on the internet can be summed up in one word: bleak’ (Wineburg et al., 2016: p. 4). While the SHEG could make this determination against reasonable expectations of what constitutes adequate or desired performance, this study benefits from the comparative value of replicating the tasks, whereby performance can be further evaluated against the existing SHEG results. The results from the students measured in Finland could be considered mixed if taken in isolation—with Task 1 and Task 5 revealing the highest outcomes at the beginning level—but when compared to the U.S., the IBDP students in Finland revealed consistently superior outcomes, at times reaching levels of mastery equivalent to the level at which the U.S. students performed at the beginning.

While the drastic differences in outcomes between the U.S. cohorts and those measured for this study are revealing, the differential in performance outcomes between the two cohorts measured separately in Finland reveal the IB2 performing at higher outcomes than the pre-IB per task, with these differences reaching levels of statistical significance when the tasks are combined. Given three possible results—that the IB2 would have performed better, the same, or worse—one would expect the cohort who have spent nearly two years in an intensive mixed infusion environment to, at least, not perform worse than the younger cohort which has yet to enter the programme. The differential in outcomes by the IB2 comparative to the pre-IB group are made more relevant when considering that the SHEG, in response to direct inquiry, claims that there were no measurable differences between grade levels in their study (Breakstone, 2018).

With over three-fourths of intervention studies examining the CT-embedded approaches of either immersion or infusion (Abrami et al., 2008; Behar-Horenstein and Niu, 2011), there remains less evidence toward their effectiveness and indeed some evidence toward their facilitating no measurable effect compared to the general and mixed approaches which facilitate CT skills development in a separate course (Tiruneh et al., 2014). While there remains work to be done in determining the extent to which any one approach may be conclusively more effective than another, the results of this study concur with previous scholarship in the field to suggest that approaches explicitly facilitating CT as a course separate from subject area integration reveal stronger outcomes than those which implicitly embed CT into subject area coursework.

Programme for the International Assessment of Adult Competencies

Although it has not received as much attention as the Programme for International Student Assessment (PISA) results, the OECD’s Programme for the International Assessment of Adult Competencies (PIAAC) includes measurements of ‘problem solving in technology-rich environments’ in addition to testing for literacy and numeracy skills. The defining parameter for this measurement is ‘the capacity to access, interpret and analyse information found, transformed and communicated in digital environments’ (OECD, 2013), with interpretation and analytical skills touching on the same CT elements found in the SHEG tasks. One advantage of considering PIAAC results, which are focused on the working-age population between ages 16 and 65, is that they provide insight into the issue of CT transfer for lifelong learning.

The first, and as of now only, PIAAC measurement took place over 2011–2012 with the results published in 2013. As with the PISA results, Finland performed well comparative to other nations, with only Japan performing higher in numeracy and literacy and only neighbouring Sweden scoring higher in problem solving within technology-rich environments. While 8.4% of Finnish adults

revealed proficiency at the highest achievement level, this is comparative to an average of 5.8% for adults in all participating countries. Further, 33.2% attained the second highest proficiency level in digital problem solving compared with the overall country average of 28.2%, and 61.9% of Finland's younger adults aged 16–24 achieved the top two levels compared with 50.7% of young adults across all participating countries. This is only 1.5% below South Korea, where young adults attained the highest scores in problem solving, and is 24.3% higher than the U.S. where young adults attained the lowest scores (2013).

International Civic and Citizenship Study

Further indications of Finland's success specific to CT skills development come from the International Civic and Citizenship Study (ICCS) conducted by the International Association for the Evaluation of Educational Achievement. When asked to prioritise three of the most important aims of civic and citizenship education, 82% of teachers polled in Finland indicated 'promoting student independent and critical thinking' as a major aim (Schulz et al., 2017: p.59). This was not only the top choice by Finnish teachers at over 25% higher than the second highest aim selected; it was also rather notably the highest percentage of any participating country (of which teachers from only two other nations selected 'independent and critical thinking' by more than 10% of the ICCS average). These data may also indicate a likelihood of subject area teachers in Finland explicitly facilitating CT skills development as per the mixed infusion model, although further research would be required to establish if this is in fact the case.

Limitations

A natural challenge to the implication that these results may indicate the IBDP fostering such development is the possibility that students of the IB2 cohort are naturally stronger performers, with other extraneous variables aside from the curriculum affecting their higher performance outcomes. This was accounted for by two metrics which help to demonstrate the similarities of the cohorts: entrance examination results for entry to the pre-IB programme, and the official IBDP scores from previous graduating classes. The entrance examinations are created locally by the school and the lowest accepted score for the IB2 cohort was 14.81 while the lowest score for the pre-IB was a very similar 14.53 (Valtanen, 2018). The average IB Diploma scores for the school in the past five years showed little variation, between 32 and 35 points (from a maximum of 45), with 95–100% of the candidates earning the Diploma. While this indicates that the school consistently performs above the IB world average of 30 points and approximately 80% Diploma pass rate (IB, 2018), it confirms that there is little by way of academic variation amongst the cohorts to explain the internal differentials.

While the two cohorts measured are academically comparable, the scholastic excellence of both participant groups comparative to other schools in Finland and other IBDP schools should be considered. In addition to Finland's already outstanding outcomes in various educational measurements, the students admitted into the IBDP at the school in the present study represent those who perform above average within a country that already produces learners who perform above international average. As such, the results between both cohorts at the IBDP school in Finland and the sampled schools in the U.S.—while at times drastic in their differences—are not quite so surprising. This invites further testing amongst more normative performing schools under various curricula to narrow the variables which appear most likely to influence the outcomes.

It is of further importance not to conflate the measurement of two separate cohorts in different stages of study from the advantages inherent to a pretest-posttest design conducted on a singular cohort over time. While a more experimental design offers advantages in terms of causal determination over this study, the establishment of the similarity of the cohorts measured helps alleviate some of the limitations associated with the causal-comparative design. Considering threats to the tasks such as language barrier, cultural bias, and differing testing conditions, the results appear to neutralise many of these concerns given the superior performance outcomes by the students in Finland over those in the U.S. Post-survey data reveal that students' digital habits and ways in which they consume online information align with the assumptions guiding the SHEG tasks: all participants in the Finland cohorts indicated that they are users of social media, with the vast majority indicating daily usage across several different networks.

Recommendations

One concrete recommendation which emerges from this study is for educational curricula of any socio-educational context to consider implementing explicit coursework in CT as a separate and compulsory component of established curricula along with core subjects such as language, literature, social studies, science, and mathematics. While further studies are required to determine the extent to which it may prove even more advantageous to explicitly embed CT into subject courses as per the mixed infusion approach, the research is clear in suggesting that curricula which implement a specific course in CT as per the general and mixed approaches reveal higher outcomes in CT skills development. Ideally, and given CT's increased importance in the development of what are often referred to as '21st century skills', CT should: 1) be heavily structured into teacher training programmes; 2) include separate licensing and certification for CT teachers; and 3) become a permanent fixture within curricula such as is found in the design and structure of the IBDP and the NCC.

In addition to increasing development of CT skills generally through explicit coursework, there appears to be a growing demand for explicitly developing CT skills specific to digital media literacy. Developments such as the adaptation in a TOK course companion to specifically address skills to determine 'fake news' from genuine media (Dombrowski, 2017), and a recent initiative in Finland to send professional journalists to schools to share their expertise on journalistic practices and social responsibility in order to help further facilitate skills in media literacy (Koponen, 2018), reveal explicit action toward further developing the skill sets in CT transfer specific to online environments. However, there remains 'a lack of comprehensive evaluation data of media literacy efforts' (Bulger and Davidson, 2018: pp. 3-4) which require further scrutiny and scholarship. The results of the present study and the existing scholarship indicate that such efforts toward explicitly facilitating these skills will continue to result in stronger development for determining the credibility and reliability of online information.

Other considerations are methodologically related to approaches to education science generally, which include advocating for more intervention studies into the general and mixed approaches to CT instruction, and for increased replication of studies across differing socio-educational environments. Regarding the former, it would greatly benefit research efforts in the field to separate the category of mixed infusion from mixed immersion for better comparative evaluation, particularly given that approaches which explicitly facilitate CT skills appear to lead to stronger outcomes. The present study provides an example of the benefits for the latter by introducing the socio-educational environment of an IBDP school in Finland as an extraneous variable from which the results

of the U.S. cohort from the SHEG study should not be generalised beyond its own socio-educational context.

Further research

With a demonstrable deficit in CT intervention studies—particularly at the upper high school level—there exists an apparent need for more research on what approaches for teaching and learning CT remain most effective. Follow-up studies which could both challenge or reinforce the initial implications of this study include replicating the materials across more average-performing Finnish upper secondary students to include at least grades 10 and 12. This would separate the IBDP and the NCC as variables of interest for influencing the higher outcomes to better consider the differential between the pre-IB and IB2. Should Finnish students in the NCC for General Upper Secondary Education overall perform similarly to the U.S. cohort, this would weaken the possibility of the Finnish curriculum and its mixed immersion approach positively affecting the outcomes. Should the Finnish students' results be equal to or better than those of the IBDP students in Finland, the IBDP and its mixed infusion approach are weakened as being considered additional influences on the stronger outcomes beyond the effectiveness of the NCC. Should the NCC for General Upper Secondary Education students perform slightly lower than the IBDP cohort, this could indicate the relative strength of the NCC's mixed immersion approach against just immersion, which is yet not as effective as the outcomes from the IBDP's mixed infusion approach.

The next logical replication would be at an average-performing IBDP school to isolate Finland as a confounding variable of interest and focus on the potential effectiveness of the IBDP and possibly by extension the mixed infusion approach. To maximise comparability with the SHEG study, a sample of IBDP students in California should be studied. As it was also determined that CT is an explicit component of the educational objectives in other countries and some U.S. states (Ennis, 2018: p.165; Silva, 2009: p. 630), the value of replicating the study to further isolate the effectiveness of explicit instruction in these states and other national curricula and socio-educational environments would provide further value to the overall implications.

Other natural extensions include replicating the materials for curricula other than the IBDP, NCC and CCSS which follow the mixed infusion or mixed immersion approaches, to test for outcome correlations amongst the approaches. This would help determine the extent to which the categorical approach can be generalised beyond its curriculum of implementation.

Acknowledgements

We would like to acknowledge Professor Erno Lehtinen for his leadership and guidance, our international colleagues and classmates at the University of Turku for providing feedback and assistance throughout the developmental stages of this research – particularly Avanti Chajed for her detailed oversight and keen eye for constructive support – and Dr Joel Breakstone of the Stanford History Education Group for granting permission to utilise their tasks and kindly providing invaluable correspondence and data for comparison.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Abrami P, Bernard R, Borokhovski E, Wade A, Surkes M, Tamim R and Zhang D (2008) Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, 78(4), 1102–1134.

- Anderson T, Howe C, Soden R, Halliday J and Low J (2001) Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science*, 29(1), 1–32.
- Australian Curriculum, Assessment and Reporting Authority (n.d.) Critical and creative thinking learning continuum. *Australian Curriculum*. Retrieved from <https://www.australiancurriculum.edu.au/media/1072/general-capabilities-creative-and-critical-thinking-learning-continuum.pdf>.
- Bee S, Black A, Cassels P, Drysdale E, Erhardt E, Free T, Grossman J, Kahn M, and Taub M (Writers), Allen A and Pennolino P (Directors) (2017, October 11) Finnish fake news [Television series segment]. In S Bee, J Jones, T Hernandez, M Kahn, A Camillo, P King (Executive Producers) and T Hall (Producer), *Full frontal with Samantha Bee*. New York: TBS.
- Behar-Horenstein LS, and Niu L (2011) Teaching critical thinking skills in higher education: A review of the literature. *Journal of College Teaching & Learning*, 8(2), 25–42.
- Bensley DA and Haynes C (1995) The acquisition of general purpose strategic knowledge for argumentation. *Teaching of Psychology*, 22(1), 41–45.
- Breakstone J (2017) *Personal Communication*, November 2, 2017.
- Breakstone J (2018) *Personal Communication*, April 19, 2018.
- Bulger M and Davison P (2018) The promises, challenges, and futures of media literacy. *Data & Society Research Institute*. Retrieved from <https://datasociety.net/output/the-promises-challenges-and-futures-of-media-literacy/>
- California Department of Education (2013, March) English language arts & literacy in history/social studies, science, and technical subjects. *California Common Core State Standards*. Retrieved from <https://www.cde.ca.gov/be/st/ss/documents/finalelaccsstandards.pdf>
- California Department of Education (2015) English language arts / English language development framework for California public schools. *Curriculum Framework and Evaluation Criteria Committee*. Retrieved from <https://www.cde.ca.gov/be/st/ss/documents/finalelaccsstandards.pdf>
- California Department of Education (2017) History–social science framework for California Public schools. *Curriculum Framework and Evaluation Criteria Committee*. Retrieved from <https://www.cde.ca.gov/ci/hs/cf/documents/hssframeworkwhole.pdf>
- Dewey J (1910) *How we think*. New York: D. C. Heath & Co.
- Dombrowski E (2017, March 27) TOK and ‘fake news’: 3 tips, 2 downloads, and 3 resources. Theory of Knowledge blog, *Oxford Education Blog*, March 27, 2017. Retrieved from <https://educationblog.oup.com/category/theory-of-knowledge>
- Domonoske C (2016, November 23) Students have 'dismaying' inability to tell fake news from real, study finds. *National Public Radio*. Retrieved from <https://www.npr.org/sections/thetwoway/2016/11/23/503129818/study-finds-students-have-dismaying-inability-to-tell-fake-news-from-real>
- Ennis R (1989) Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, 18(3), 4–10.
- Ennis R (1991) Critical thinking: A streamlined conception. *Teaching Philosophy*, 14(1), 4–18.
- Ennis R (2018) Critical thinking across the curriculum: A vision. *Topoi*, 37(1), 165–184.
- Facione P (1990) Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. *The Delphi Report*. Millbrae, CA: California Academic Press.
- Fillion N and Martelli D (2017, November 10) Overview of critical thinking in the BC K-12 curriculum revisions. *Simon Fraser University*, British Columbia, Canada. Retrieved from https://www.sfu.ca/content/sfu/conferences/criticalthinking/positionpaper/_jcr_content/main_content/download/file.res/Critical%20Thinking%20position%20paper.pdf
- Finnish National Board of Education (2016) *National core curriculum for basic education 2014*. Helsinki: Author.
- Finnish National Board of Education (2016) *National core curriculum for general upper secondary schools 2015*. Helsinki: Author.
- Gewertz C (2008, October 13) States press ahead on 21st century skills. *Education Week*, 28(8), 21–23.
- Glaser E (1941) *An experiment in the development of critical thinking*. New York: Bureau of Publications, Teachers College, Columbia University.
- Glaser R (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39(2), 93–104.

- Glevey K (2008) Thinking skills in England's National Curriculum. *Improving Schools*, 11(2), 115–125.
- Halpern D (1998) Teaching critical thinking for transfer across domains. Dispositions, skills, structure training, and metacognitive monitoring. *The American Psychologist*, 53(4), 449–455.
- Hobbs R and Jensen A (2009) The past, present and future of media literacy education. *Journal of Media Literacy Education*, 1(1), 1–11.
- International Baccalaureate (2017) The IB Diploma Programme statistical bulletin, May 2017 examination session. *IB Global Centre*, Cardiff. Retrieved from <http://www.ibo.org/contentassets/bc850970f4e54b87828f83c7976a4db6/dp-statistical-bulletin-may-2017-en.pdf>
- Koponen H (2018, January 24) New Finnish project brings journalists to schools to teach media literacy. *International Press Institute*. Retrieved from <https://ipi.media/new-finnish-project-brings-journalists-to-schools-to-teach-media-literacy/>
- Krippendorff K (2018) *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: SAGE.
- Larsson K (2017) Understanding and teaching critical thinking—A new approach. *International Journal of Educational Research*, 84, 32–42.
- Leen C, Hong H, Kwan F and Ying T (2014) Creative and critical thinking in Singapore schools. *Office of Education Research, National Institute of Education, Nanyang Technological University*. Retrieved from https://www.nie.edu.sg/docs/default-source/nie-working-papers/niewp2_final-for-web_v2.pdf?sfvrsn=2
- Makel M and Plucker J (2014) Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 304–316.
- Mayer R (1992) Teaching of thinking skills in the sciences and mathematics. In D. Halpern (Ed.), *Enhancing Thinking Skills in the Sciences and Mathematics* (pp. 95–116). New Jersey: Lawrence Erlbaum Associates.
- McGrew S, Breakstone J, Ortega T, Smith M and Wineburg S (2018) Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, 46(2), 165–193.
- McGrew S, Ortega T, Breakstone J, and Wineburg S (2017) The challenge that's bigger than fake news: Civic reasoning in a social media environment. *American Educator*. Retrieved from <https://www.aft.org/sites/files/periodicals/fall2017mcgrew.pdf>
- McPeck J (1990) Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher*, 19(4), 10–12.
- National Association for Media Literacy Education (2007) *Core principles of media literacy education in the United States*. Retrieved from <https://namle.net/publications/coreprinciples>
- National Council for the Social Studies (2016) Media literacy. *Social Education*, 80, 183–185.
- Nickerson R (1988) On improving thinking through instruction. *Review of Research in Education*, 15(3), 3–57.
- Ontario Ministry of Education (2015) The Ontario curriculum grades 11 and 12. *Canadian and World Studies*. Toronto, ON: Queen's Printer for Ontario. Retrieved from www.edu.gov.on.ca/eng/curriculum/secondary/2015cws11and12.pdf
- Organisation for Economic Cooperation and Development (2013) Country note: Finland. *Survey of adult skills first results*. Retrieved from <http://www.oecd.org/skills/piaac/Country%20note%20-%20Finland.pdf>
- Orszag A (2015) Exploring Finnish university students' perceived level of critical thinking. (Master's thesis). *University of Jyväskylä, Jyväskylä, Finland*.
- Paul R (1995) *Critical thinking: What every person needs to survive in a rapidly changing world*. Santa Rosa, CA: Foundation for Critical Thinking.
- Perkins D and Salomon G (1988) Teaching for transfer. *Educational Leadership*, 46(1), 22–32.
- Plath D, English B, Connors L and Beveridge A (1999) Evaluating the outcomes of intensive critical thinking instruction for social work students. *Social Work Education: The International Journal*, 18(2), 207–217.
- Resnick LB (1987) *Education and Learning to Think*. Washington DC: National Academy Press.
- Saiz C and Nieto A (2010) Critical thinking: A question of aptitude and attitude? *Inquiry: Critical Thinking Across the Disciplines*, 25(2), 19–26.

- Schulten K and Brown A (2017) Evaluating sources in a ‘post-truth’ world: Ideas for teaching and learning about fake news. *The New York Times*. Retrieved from <https://www.nytimes.com/2017/01/19/learning/lesson-plans/evaluating-sources-in-a-post-truth-world-ideas-for-teaching-and-learning-about-fake-news.htm>
- Schulz W, Ainley J, Fraillon J, Losito B, Agrusti G and Friedman T (2017) Becoming citizens in a changing world: International civic and citizenship education study 2016 international report. *International Association for the Evaluation of Educational Achievement*. Retrieved from https://iccs.iea.nl/fileadmin/user_upload/Editor_Group/Downloads/ICCS_2016_International_report/pdf
- Shellenbarger S (2016, November 21) Most students don’t know when news is fake, Stanford study finds. *The Wall Street Journal*. Retrieved from <https://www.wsj.com/articles/most-students-dont-know-when-news-is-fake-stanford-study-finds-1479752576>
- Silva E (2009, May 1) Measuring skills for 21st-century learning. *Phi Delta Kappan*, 90(9), 630–634.
- Simpson J, Weiner E and Oxford University Press (2018) *The Oxford English Dictionary*. Oxford: Clarendon Press.
- Standish R (2017, March 1) Why is Finland able to fend off Putin’s information war? *Foreign Policy*. Retrieved from <http://foreignpolicy.com/2017/03/01/why-is-finland-able-to-fend-off-putins-information-war/>
- Tiruneh D, Verburch A and Elen J (2014) Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies* 4(1), 1–17.
- Valtanen J (2018) *Personal Communication*, February 26, 2018
- Wineburg S, McGrew S, Breakstone J and Ortega T (2016) Evaluating Information: The Cornerstone of Civic Online Reasoning. *Stanford Digital Repository*. Retrieved from <http://purl.stanford.edu/fv751yt5934>
- Wineburg S and McGrew S (2016, December 13) Most teens can’t tell fake from real news. *PBS News Hour*. Retrieved from <https://www.pbs.org/newshour/education/column-students-cant-google-way-truth>

Shane Horn taught courses in the social sciences, humanities, literary studies and performing arts at the upper secondary level across international schools in China, Nepal, and Kazakhstan for the past decade before pursuing graduate studies at the University of Turku in Finland. From September 2018 he has been the International Baccalaureate Diploma Theory of Knowledge coordinator and teacher of Literature and Performance, Theatre, and Global Politics at Amman Baccalaureate School in Jordan.

Koen Veermans is a researcher in the Department of Teacher Education at the University of Turku in Finland. He obtained his PhD on intelligent support for discovery learning from the University of Twente in the Netherlands on research in the context of an authoring system for technology-enhanced inquiry learning. In Finland he has continued research on technology-enhanced inquiry learning in the context of Academy of Finland projects and EU funded projects, and is academic director of the EdLearn International Masters programme.