

Critical thinking in national tests across four subjects in Swedish compulsory school

Thomas Nygren ^a, Jesper Haglund ^b, Christopher Robin Samuelsson ^b,
Åsa Af Geijerstam ^a and Johan Prytz ^a

^aDepartment of Education, Uppsala University, Uppsala, Sweden; ^bDepartment of Physics and Astronomy, Uppsala University, Uppsala, Sweden

ABSTRACT

Critical thinking is brought to the fore as a central competence in today's society and in school curricula, but what may be emphasised as a general skill may also differ across school subjects. Using a mixed methods approach we identify general formulations regarding critical thinking in the Swedish curriculum of school year nine and seven more subject-specific categories of critical thinking in the syllabi and national tests in history, physics, mathematics and Swedish. By analysing 76 individual students' critical thinking as expressed in national tests we find that a student that thinks critically in one subjects does not necessarily do so in other subjects. We find that students' grades in different subjects are closely linked to their abilities to answer questions designed to test critical thinking in the subjects. We also find that the same formulations of critical thinking in two subjects may mean very different things when translated into assessments. Our findings suggest that critical thinking among students comprise different, subject-specific skills. The complexity of our findings highlights a need for future research to help clarify to students and researchers what it means to think critically in school.

KEYWORDS

Critical thinking; history; mathematics; Swedish; physics; national tests; mixed method; explorative factor analysis

Introduction

Critical thinking today is underscored as a pivotal skill in society and in curricula. The ambiguous concept of critical thinking is emphasised as a central twenty-first century skill – central in education, work life and civil society (Binkley et al., 2012; European Commission, 2016). However, critical thinking is a complicated skill which has proven difficult to promote in teaching, perhaps due to the fact that critical thinking must be understood as more than a single intellectual and academic skill (Willingham, 2008). Still, when a student leaves compulsory school in Sweden, he or she should, according to the overarching goals of the curriculum, be able to in all subjects “make use of critical thinking and independently formulate standpoints based on knowledge and ethical considerations” (Skolverket, 2011, p. 15). Furthermore, in the syllabi, critical thinking is interpreted and operationalised in different subject-specific ways (Skolverket, 2011,

CONTACT Thomas Nygren  thomas.nygren@edu.uu.se  Department of Education, Uppsala University, Box 2136, Uppsala750 02, Sweden

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

see further details below). This means that critical thinking is formulated as a general ability and also in more specific ways depending on what subject is to be studied.

Thus, today students come into the classrooms and are asked to think critically in different subjects, and we as researchers in education should be able to say what this means in school subjects and across disciplinary boundaries. At this point, however, research has not managed to establish to what degree critical thinking is a matter of distinct skills in different school subjects and to what degree it is primarily a general ability.

Aim and research questions

The aim of this study is to describe and discuss the relationship between students' critical thinking as expressed in syllabi and national tests in, and between, school subjects in school year nine. Through the study of students' critical thinking in mathematics, physics, history, and Swedish, as operationalized in national tests, we investigate how available empirical data can contribute to the understanding of critical thinking as a general or subject-specific skill in school.

In this study, we address the following research questions, in relation to how critical thinking is expressed in Swedish compulsory school and national tests:

- How is critical thinking expressed and operationalised in different school subjects?
- To what degree is critical thinking a subject-specific or general skill?
- To what degree is critical thinking a skill that is distinct from general subject knowledge?

Using individual students' test scores and grades in and across disciplines makes it possible to analyse critical thinking as primarily a general or subject-specific skill. We hypothesise that if critical thinking is a general skill there should be a good correlation across categories of critical thinking in the test scores in the different subjects. If critical thinking is primarily a general skill, then a cluster analysis of students' test scores of critical thinking in different subjects should give a single general cluster or an even distribution across subjects. In contrast, if critical thinking is primarily a subject-specific skill then students may have a high score in critical thinking in one subject and a low score in another. Correlations across categories of subject-specific critical thinking should in this case be low and a cluster analysis would in this case show clusters of students' results in clearly separated subject-specific groups.

Previous research

Research on critical thinking is today noted as a field with a multitude of perspectives and contrasting views (Abrami et al., 2015). General approaches to teaching critical thinking could involve topics such as logic and assessing the sender of a message, without connection to a particular subject. In contrast, subject-specific approaches are based on the premise that critical thinking can only be developed in relation to some kind of content, in terms of explicit emphasis on critical thinking in subject teaching, or the assumption that critical thinking develops implicitly with deeper subject knowledge. To what extent critical thinking is a subject-specific skill or more of a general skill has

been discussed since the 1980s, especially in the USA (Ennis, 1985a; Facione, 1990; McPeck, 1981; Perkins & Salomon, 1989; Siegel, 2013). This is an ongoing debate (Davies, 2013; Moore, 2011) and current reviews of research note how, especially in higher education, there are still contrasting perspectives on critical thinking as primarily a matter of generic or content-specific skills (Abrami et al., 2015). Some researchers have found that critical thinking as a generic skill may be supported by special courses in critical thinking (Royalty, 1995; Sá, Stanovich, & West, 1999) or small interventions (Solon, 2007), while other researchers underline the importance of stimulating critical thinking in disciplinary ways paying attention to the content and context (Halliday, 2000; Moore, 2011; Smith, 2002).

Subject-specific educational research has shown differences in critical thinking between experts and novices within different subjects (Shanahan, Shanahan, & Misischia, 2011; Wineburg, 2001), and as part of the expanding research on disciplinary literacy, there are several studies of how students express critical thinking in different subjects (Lundqvist, Säljö, & Östman, 2013; Pearson, Moje, & Greenleaf, 2010; Shanahan & Shanahan, 2012; Stevens, Wineburg, Herrenkohl, & Bell, 2005). However, there are very few comparative studies across subjects, as noted by Herrenkohl and Cornelius (2013). Even though the problem of transfer from one learning situation to another has been noted (Ennis, 1989; Perkins & Salomon, 1989), there is a striking lack of studies following individual students' learning between subjects. Research in Sweden has studied critical thinking in the social sciences (Broman, 2009), Swedish (Wyndhamn, 2013), history (Nygren, Sandberg, & Vikström, 2014; Rosenlund, 2016), and philosophy (Hjort, 2014) in separate studies or across subjects in different groups, but not following the same students' abilities across disciplinary boundaries. Although some studies show correlation between grades in different subjects (e.g. Stenhag, 2010) no empirical study yet has analysed patterns with specific regard to critical thinking.

The lack of research across disciplines may be explained by academic traditions and cultures of separating subjects in school settings. For example, the extensive research of disciplinary literacy is today criticised for taking subject boundaries for granted and a lack of attention to relations across disciplinary boundaries or ignoring disciplinary perspectives in favour of more general didactic perspectives (Stevens et al., 2005). Even the current large scale OECD (2015a,b) programmes to assess “creative and critical thinking skills in education” lack a cross-disciplinary perspective that pays sufficient attention to individuals in different subject-specific settings. The question of whether or not critical thinking among students is a primarily general or subject-specific skill therefore remains unanswered.

Theoretical consideration

The ambiguous concept of critical thinking has previously been defined in numerous ways. In this study, we use a definition by experts of a Delphi panel, organised by the American Philosophical Association, which defined critical thinking “to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based” (Facione, 1990, p.

3). We find this definition to be inclusive and clear enough to capture critical thinking, whether interpreted as a generic or subject-specific skill. This definition by Facione (1990) is widely used today in research on critical thinking in education, for instance to conduct meta-analyses of critical thinking as both generic and subject specific (Abrami et al., 2015). We find that the definition includes formulations open for a more generic perspective *and* a more subject-specific perspective on critical thinking, without prioritising one before the other.

Current perspectives on critical thinking in education often date back to the theoretical debate between Robert Ennis and John E. McPeck. They were prominent in the development of theories regarding critical thinking, as they discussed different definitions of and perspectives on critical thinking. Ennis (1962, 1985a) emphasised general aspects of critical thinking which students can learn in courses separate from academic subjects. Critical thinking in this general sense can, according to Ennis, also be tested in assessments as general-thinking skills (1985b). In contrast McPeck argued that critical thinking always relates to subject content and contexts. Underscoring the subject-specificity of critical thinking, McPeck (1981, p. 7) stated that critical thinking is always a matter of “reflective scepticism within the problem area under consideration.” From this perspective, it is evident how beliefs, grounds of arguments and conclusions may differ depending on content, context and disciplinary traditions. McPeck (1990b, p. 21) contended that “if we improve the quality of understanding through the disciplines. . . you will then get a concomitant improvement in thinking capacity.” In line with this perspective, advocates of disciplinary literacy find that critical thinking may hold different qualities between subjects (Moore, 2011; Shanahan et al., 2011).

In a head-to-head debate in *Educational Researcher*, Ennis (1989) and McPeck (1990a) made clear how they had different theoretical perspectives on critical thinking. What this debate highlighted was that they both, in a limited way, acknowledged the position of the opponent. Ennis (1989, p. 8) found that there might be “interfield variations” between disciplines and McPeck (1990b, p. 12) said that “[f]or what it is worth, I believe there are, in fact, some very limited general thinking skills”. Thus, in theory it is not a clear-cut case of general or subject-specific skills. And Ennis (1989, p. 9) suggested an empirical “examination of the degree of commonality of the critical thinking aspects found in the different standard existing disciplines and school subjects” to better understand the crucial and confusing question of subject specificity. This is what we do in this study.

Materials and methods

In order to analyse how critical thinking is expressed and operationalised in different school subjects and empirically test if it is primarily a general or subject-specific skill, we used the Swedish curriculum and syllabi of physics, history, mathematics, and Swedish, and national tests at school year nine as the main sources of material. Swedish national tests and test rubrics are developed and validated centrally, and the classes’ regular teachers – typically in collaboration with their teacher colleagues – are marking the tests and reporting on the results. Grading guidelines, formulated on a national level by groups of test experts, are extensive, detailed and especially focused on students’ skills in different subjects. The selection of subjects was based upon the fact

that they may be noted as different with regard to disciplinary literacy and they are all included in national tests (Goldman et al., 2016; Shanahan et al., 2011).

With a theory and problem-driven mixed methods approach, we closely studied formulations of syllabi and questions in the national tests, and on an aggregated level we analysed students' test results within the analysed discourses (Johnson & Onwuegbuzie, 2004; Namey, Guest, Thairu, & Johnson, 2008; Tashakkori & Teddlie, 2010). The quantitative method makes generalisations possible, while the qualitative analyses give valuable interpretations of the statistics and provide nuanced links to critical thinking as formulated in theory, syllabi and national tests. Syllabi and national test questions were analysed in light of theories of critical thinking and disciplinary literacy. In a close reading of syllabi and national test questions, we found different constructions of critical thinking in the syllabi and in the questions prompting students to think critically in the national test. Based on this analysis, we selected the items that were most consistent with Facione's definition to include in our statistical analysis of students' critical thinking across subjects. The selected items that addressed critical thinking typically required students to provide quite long answers in coherent paragraphs, as opposed, for example, to multiple-choice questions.

To empirically test whether a student noted for critical thinking in history was also noted for this in mathematics, physics, and Swedish, and vice versa we studied individual students' national tests results within and across the different subjects. We analysed graded answers from 76 students (15–16 years old) who had taken tests in all four subjects. The national tests, questions, guidelines and answers from 2013 are available as examples for teaching and research and thus constituted our empirical material. This analysis was conducted as a collaboration of Thomas Nygren, Jesper Haglund, Åsa af Geijerstam, and Johan Prytz with expertise from the fields of history, physics, Swedish and mathematics education to safeguard disciplinary perspectives in the analysis, and Christopher Robin Samuelsson with expertise in statistics. Collaboration in the analysis also supported interpretations to be valid and reliable outside the subjective perspective of a single researcher (Krippendorff, 2013).

All test results come from a public school in a suburban part of a Swedish city with ca. 100,000 inhabitants. From this school in 2013, 78.9 percent of the students received grades to grant them access to upper secondary schooling; this is 8.7 percent units lower than the national average of 87.6. Of 89 students listed in school year nine, 76 wrote national tests in all four subjects. Especially, students with low grades did not take all the tests. The students in this study hold a wide range of results and grades and the final-grade scores among the students range between 67.5 and 315 with an average of 213 on a national scale from 0 to 320.

As described in the subsequent Results section, a total of seven categories of critical thinking were identified in the national tests, two categories each in the subjects of Swedish, history, physics, and one category in mathematics. Individual test items in the tests that reflected these categories were identified. In the quantitative analysis of the data, the original data with grades F, E, C and A, set by the teachers, was translated into numerical values in line with common Swedish recalculation of grades on four defined grade levels (F = 0, E = 10, C = 15, A = 20). The scores in the seven categories of critical thinking constitute seven variables in our analysis. These variables were initially processed and investigated in a visual pattern analysis. Here we used colour scales to

visually make evident individual students' scores across test items of critical thinking across the subjects. This was followed by a statistical analysis also including the students' final grades in the subjects Swedish, history, physics and mathematics (graded F, E, D, C, B, or A). Grades translated into numerical values (0, 10, 12.5, 15, 17.5, or 20) constituted four additional variables, interpreted as indications of the students' more general subject knowledge.

The numerical values used could either be viewed as ordinal or interval. Interval scales requires the data to have equal differences between subsequent data points (Everitt, 2005). In our case, the longer distance between 0 (grade F) and 10 (grade E) may be considered an exception, but could cause problems. To be considered ordinal, the data need to be ordered (as grades are) but do not need to have equal distances between the data points. For data with an interval scale, Pearson's correlation can be used to calculate correlations between variables. For ordinal data, Spearman's correlation can be used. In our case, Spearman's correlations were calculated between all variables, including the categories of critical thinking and the students' final grades, since different scales were compared. In addition, Pearson's correlations were calculated between the critical thinking variables, since the values are close to equal differences.

In addition, the student scores on the test items were also analysed through factor analysis with maximum likelihood (FML) and principal factor analysis (PFA). Factor analysis is a method that is suitable for exploring patterns in data based on correlations or covariance between variables. In this case, a correlation matrix based on the Pearson method and an interval scale for the data was used. Using 76 data points in 7 variables is within the recommended span for factor analysis, but on the lower end, since recommendations often state 50 points as a lower limit to make it possible to find coherent correlations between the variables (Everitt, 2005; Nakazawa, 2011; Zhao, 2009). Measures of sampling adequacy regarding partial correlation evaluation of the data using a Kaiser-Meyer-Olkins sampling adequacy (KMO) test showed how the data was useful as a basis for factor analysis. KMO from 0.75 to 0.85 were well above the recommended limit of 0.5 (Hair, Anderson, Babin, & Black, 2010; Nakazawa, 2011). The sample size is not the ideal ratio of $\sim 20:1$ but $\sim >10:1$ can still be considered a reasonable ratio. Using FML assumes a normal distribution of the data. In our case, the data may not be normally distributed due to the criteria-based grading system in Sweden. To take this fact into account, PFA, which does not require any assumption of the distribution, was used to complement the FML approach. FML and PFA create clusters of data useful in this study to identify if students' individual test scores are organised in a general way or separated in more subject-specific ways.

Results

Bearing in mind the definition of critical thinking as "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (Facione, 1990, p. 3) we found that the compulsory national curriculum for school year nine holds a number of formulations regarding critical thinking.

Identification of aspects of critical thinking in different subjects

In the syllabus for history we found two separate categories related to critical thinking. The syllabus states that students should “critically examine, interpret and evaluate sources as a basis for creating historical knowledge,” (Skolverket, 2011, p. 163) this cognitive act of critically scrutinising information and judging it with regard to the source, corroboration of evidence and the context has been labelled historical thinking. Historical thinking in this category is closely related to the academic practices of reading like a historian (Wineburg, 1991, Wineburg, 2001). We call this category of critical thinking *critical historical thinking*. In the second category of formulations, we found formulations stating that teaching should make students “reflect over their own and other’s use of history in different contexts and from different perspectives” (Skolverket, 2011, p. 163). This critical thinking skill is directed towards judging different uses and misuses of history and what has been labelled historical culture (Karlsson, 2014) and abilities of multiperspectivity or having a historiographic gaze (Nygren, Vinterek, Thorp, & Taylor, 2017). In this study, we call this *critical thinking about uses of history*. This category may involve interpretations about present events which go outside the academic ideal of critical thinking in history, but in this study, it is a useful category since it may test a separate critical thinking skill.

The fact that the national test questions have been designed in line with theories of historical thinking and uses of history has previously been noted (Eliasson, Alvé, Yngvéus, & Rosenlund, 2015; Samuelsson & Wendell, 2016) and our analysis confirms this. In the national test, we found a number of questions designed to assess students’ *critical historical thinking* skills, described in the test as abilities to “critically examine, interpret and evaluate sources” (Skolverket, 2013a). Students were for instance asked to review and use primary sources about slave trade in the eighteenth century, women’s rights during the French revolution, and the social history of the textile industry in the mid-twentieth century. Other questions were designed to assess students reasoning skills regarding “how history has been used and can be used” and how current perspectives on society may be influenced by “divergent perspectives on the past,” what we call *critical thinking about uses of history*. Here students were asked to analyze contemporary uses of history by Save the Children and the organisers of the 2012 Olympics. Students were also asked to review how Olof Palme used history to argue against bombings in Vietnam in 1972.

In the syllabus of Swedish, we found several formulations that could be interpreted as aspects of critical thinking. Students are supposed to “develop their knowledge of how to search for and critically evaluate information from various sources” (Skolverket, 2011, p. 211). Students should also learn to “read and analyse literature and other texts for different purposes” (ibid) which could include taking a critical stance in reading. Attention is also directed to different aspects of language form as students should learn e.g. “the importance of language in exercising influence” and “words and terms, their shades of meaning and value connotations”. To be able to pay attention to form aspects in language is important in critical thinking. In several of the criteria formulated for grading in Swedish, aspects of critical thinking is foregrounded. Students should learn to process information critically by “reasoning about the work and how it is related to its creator” and by “informed reasoning about the explicit and implicit messages in

different works” (Skolverket, 2011, p. 220). Also, the criteria for grading states that a student should be “asking questions and expressing opinions with well-developed and well informed arguments” and “search for, select and compile information from a varied range of sources and then apply well developed and well informed reasoning to the credibility and relevance of their sources and information” (Skolverket, 2011, p. 221). In sum, the curriculum of Swedish puts forward many aspects crucial for critical thinking where applications such as verbal reasoning, argument analysis and the critical use of sources are foregrounded.

When looking into the national test in Swedish where reading comprehension was in focus, we found that two types of critical thinking was tested. The questions were organised to test different “reading processes” (as established in the international reading test PIRLS, Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009), and we identified questions relating to aspects of critical thinking found within two of these reading processes; “interpret and integrate ideas and information” and “examine and evaluate content and textual elements” (Skolverket, 2013b). We identified a number of questions designed to assess students’ abilities to combine and interpret different types of information in a text. Students were for instance asked to interpret why a character behaves in a certain way in a literary text, formulate an opinion based on what is expressed in a text, or discuss a quotation from an informational text in two opposing ways. In this study, we call this *critical thinking in interpreting information in Swedish*. Questions in the test designed to test students’ abilities to evaluate information were fewer, but included a question asking students to find an expression for irony in a narrative text and a question asking for genre knowledge. We call this *critical thinking in evaluating content and form in Swedish*. We thus found that the syllabus of Swedish express several aspects of critical thinking that are not operationalized and tested in the reading comprehension parts of the national tests.

In our close reading of the syllabus for physics education, we found how critical thinking was formulated in two ways, critical thinking regarding physics in society and critical thinking in the science of physics. The syllabus for physics sets forth the importance of “critical examination of information and arguments which students meet in sources and social discussions related to physics” (Skolverket, 2011, p. 124). Students are moreover supposed to problematize issues regarding “energy, technology, the environment and society, and differentiate facts from values” (Skolverket, 2011, p. 127). We interpret this as *critical thinking regarding physics in society*, much in line with what Roberts (2007) has called Vision 2 in science education. This way of thinking critically should work in parallel with a critical thinking with a more internal focus on science in its own right, aligning with Roberts’ (2007) Vision 1. Here students should develop critical thinking to evaluate and judge different methods and results as part of conducting a systematic investigation, and draw conclusions from comparisons in light of theories and models developed within the science. This relates closely to the view of critical thinking in physics as “the ability to make decisions based on data” (Holmes, Wieman, & Bonn, 2015, p. 11,199). We call this subject-specific category of critical thinking *critical thinking in the science of physics*.

In line with the formulations, we found questions testing these abilities in the national test. *Critical thinking regarding physics in society* was assessed as an ability to “examine information, communicate and take a view on questions concerning energy,

technology, the environment and society” with a task where students were asked to provide arguments for a recommendation to the Swedish minister of energy of what energy sources to use for electricity production. *Critical thinking in the science of physics* was assessed as an ability to “carry out systematic investigation in physics”. Here, students were asked to plan, conduct, and draw conclusions from an experiment, and suggest ways in which the result could be made more reliable. Students were also asked to assess, from given information, which one of a house with walls of wood and a house with brick walls would heat up faster in the sun (Skolverket, 2013c).

According to the mathematics syllabus, the students should learn “to develop their knowledge in order to formulate and solve problems, and also reflect over and evaluate selected strategies, methods, models and results” (Skolverket, 2011, p. 59). In particular, the part about evaluation is in line with Facione’s definition of critical thinking. To distinguish tasks that require critical thinking in mathematics, we used Lithner’s (2008) definition of *Creative mathematically founded reasoning* (CMR). CMR involves a critical or evaluative element since considerations about plausibility is essential in CMR: it includes “arguments supporting the strategy choice and/or strategy implementation motivating why the conclusions are true or plausible.” This is in line with Facione’s definition of critical thinking. However, Lithner adds: “Mathematical foundation. The arguments are anchored in intrinsic mathematical properties of the components involved in the reasoning” (p. 266).

Following Lithner’s definition, one task in the national exams was singled out. It was deemed suitable since it clearly asked the students to engage in *critical thinking in mathematics*. The task was called the *swimming hall* (*simhallen*) and concerned three models for payment (Skolverket, 2013d). All models were linear and involved two variables: costs and number of visits. The models were presented visually as straight lines in a two-dimensional coordinate system. Apart from reading the diagram and formulating algebraic expression for each model, the students were asked to choose one of the models and account for advantages and disadvantages of all three. They then had to engage in evaluative reasoning that involved properties of mathematical concepts, for instance, relations between two variables, the inclinations of the straight lines and what it means when two straight lines cross. They were also asked to explain why the cost was proportional with the numbers of visits in some of the models. This means that the students had to motivate a conclusion on the basis of a definition of proportionality.

Student scores on critical thinking in national tests

Thus, critical thinking was formulated explicitly and implicitly in syllabi and tests. In sum, we found seven categories of critical thinking which could be used to separate subject-specific critical thinking skills from general skills, namely: (1) critical thinking in interpreting information in Swedish, (2) critical thinking in evaluating content and form in Swedish, (3) critical thinking regarding physics in society, (4) critical thinking in the science of physics, (5) critical thinking in mathematics, (6) critical historical thinking, (7) critical thinking about uses of history.

We hypothesised that if critical thinking is primarily a general skill then there would be high correlations between individual students’ scores in the different national tests on questions testing critical thinking. Our first analysis of students’ scores in the seven

categories, when translated into numerical values and sorted in descending order, showed a pattern where 17 students had similar grades on critical thinking across the categories, for instance students with only the two highest or the two lowest grades had this across all categories. Meanwhile, we also found a number of students who scored high in one category but low in another. Some of the categories stand out as more challenging for the students than others, e.g. *critical thinking in evaluating content and form in Swedish*, 16 students with the lowest score in this category had the highest score in another category. Excluding this category, we still found two students with the lowest and highest grade in different subject-specific critical thinking categories. One of them had a low score in the category of *critical thinking regarding physics in society*, but got a top score in *critical historical thinking* and *critical thinking in evaluating content and form in Swedish* as well as high scores across the other categories of critical thinking. The other students had low scores in both of the Swedish subject categories, but got a high score in critical thinking in physics, mathematics and history. We also found four students who managed very well in Swedish to “interpret information” but failed to “evaluate content and form” in combination with a mix of good and passing scores in the other categories.

Correlation analysis of student scores in national tests and final grades

A Spearman correlation matrix was then calculated (see Table 1 and Figure 1).

In the correlation matrix in Figure 1, an overall pattern is that there are positive, but moderate correlations between variables V1-V7, representing the test scores on the different critical thinking categories, both within and across the four subjects. For

Table 1. Variables used in the statistical analysis.

Variable	Category
V1	Critical thinking in interpreting information in Swedish
V2	Critical thinking in evaluating content and form in Swedish
V3	Critical thinking regarding physics in society
V4	Critical thinking in the science of physics
V5	Critical thinking in mathematics
V6	Critical historical thinking
V7	Critical thinking about uses of history
V8	Final grade in Swedish
V9	Final grade in physics
V10	Final grade in mathematics
V11	Final grade in history

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V1	1.0000000	0.3575870	0.3583033	0.1805752	0.4577861	0.3509941	0.4643477	0.6440539	0.4084638	0.4471123	0.4953869
V2	0.3575870	1.0000000	0.2353491	0.2200464	0.3876629	0.3626552	0.3975336	0.4117724	0.3314330	0.4086533	0.3278579
V3	0.3583033	0.2353491	1.0000000	0.5363891	0.6156115	0.3614002	0.5332334	0.4254548	0.6425098	0.5969431	0.5059665
V4	0.1805752	0.2200464	0.5363891	1.0000000	0.4522998	0.5107553	0.3901202	0.2938088	0.6278709	0.5842443	0.4829815
V5	0.4577861	0.3876629	0.6156115	0.4522998	1.0000000	0.4240409	0.4933762	0.5674984	0.7455721	0.7449577	0.6045624
V6	0.3509941	0.3626552	0.3614002	0.5107553	0.4240409	1.0000000	0.5018972	0.5910714	0.6066660	0.6251488	0.7040294
V7	0.4643477	0.3975336	0.5332334	0.3901202	0.4933762	0.5018972	1.0000000	0.6606687	0.6073880	0.5632614	0.6507347
V8	0.6440539	0.4117724	0.4254548	0.2938088	0.5674984	0.5910714	0.6606687	1.0000000	0.6247308	0.6006694	0.6915950
V9	0.4084638	0.3314330	0.6425098	0.6278709	0.7455721	0.6066660	0.6073880	0.6247308	1.0000000	0.8764209	0.7533463
V10	0.4471123	0.4086533	0.5969431	0.5842443	0.7449577	0.6251488	0.5632614	0.6006694	0.8764209	1.0000000	0.6978158
V11	0.4953869	0.3278579	0.5059665	0.4829815	0.6045624	0.7040294	0.6507347	0.6915950	0.7533463	0.6978158	1.0000000

Figure 1. Spearman correlation matrix of the variables representing critical thinking and final grades in the four subjects

example, the correlation between the score on *critical thinking in the science of physics* (V4) and *critical thinking in interpreting information in Swedish* (V1) is 0.18, a low, yet positive value. Also the highest correlation among the variables V1-V7, 0.62, between *critical thinking regarding physics in society* (V3) and *critical thinking in mathematics* (V5), is moderate.

In contrast, the correlations between the final grades in the involved subjects, variables V8-V11, were high compared to the rest of the correlations. For example, the correlation between the final grade in physics (V9) and mathematics (V10), two closely associated subjects, is 0.88. It is maybe more surprising that the final grade in history (V11) is also strongly correlated to that of physics (V9), with a value of 0.75.

Correlations between scores of critical thinking items and the corresponding subject grades are also rather high overall. For example, the correlations between, on the one hand, the score on critical thinking in the science of physics (V4) and critical thinking regarding physics in society (V3), and on the other, the final grade in physics (V9), are 0.64 and 0.63, respectively. The correlations between scores of critical thinking with final grades in other subjects are generally lower. For example, critical thinking in the science of physics (V4) has the following correlations with the final grades in the other subjects: Swedish (V8), 0.29; History (V11): 0.48; Mathematics (V10): 0.58.

Exceptions to these patterns are that the correlations between the scores on critical thinking in Swedish (V1 and V2) and other variables (apart from between V1 and V8, the final grades in Swedish) are quite low. In addition, there are high correlations between the score on *critical thinking in mathematics* (V5) and the final grade in physics (V9) (comparable to the correlation between the score on V5 and the grade in mathematics), 0.75, and between the score on *critical thinking about uses of history* (V7) and the final grade in Swedish (V8), 0.66.

Factor analysis of student scores on critical thinking

In addition to the correlation analysis, an explorative factor analysis was done on the Pearson correlation matrix of the student scores on the categories representing critical thinking, using maximum likelihood (FML) and principal factor analysis (PFA). Using Pearson correlations, the data was assumed to be interval-based. As seen in [Figure 2](#), positive but weak correlations were found between the variables, similar to the Spearman correlations presented above (see [Figure 1](#)). This means that the resulting factor model will explain these weak correlations, but not other variance that is not shared between the variables.

	V1	V2	V3	V4	V5	V6	V7
V1	1.0000000	0.3563417	0.3435418	0.1938364	0.4460365	0.3906727	0.4749929
V2	0.3563417	1.0000000	0.1720974	0.1988433	0.3677633	0.3602967	0.3795700
V3	0.3435418	0.1720974	1.0000000	0.5051202	0.5575712	0.3234061	0.4926310
V4	0.1938364	0.1988433	0.5051202	1.0000000	0.4151968	0.4557884	0.3300196
V5	0.4460365	0.3677633	0.5575712	0.4151968	1.0000000	0.4255617	0.4655262
V6	0.3906727	0.3602967	0.3234061	0.4557884	0.4255617	1.0000000	0.4373505
V7	0.4749929	0.3795700	0.4926310	0.3300196	0.4655262	0.4373505	1.0000000

Figure 2. Pearson correlation matrix of the variables representing critical thinking

Based on the Pearson correlation matrix, maximum likelihood estimation (FML) assuming a one-factor model yielded the following result (see [Figure 3](#)).

The loadings of all seven variables on the factor are positive and quite high. This means that the correlations between the variables can be explained by this factor to a high degree. However, the high values of the uniqueness for the variables and the low value of the proportion of variance indicate that most of the variance is unique to the variables, and cannot be explained by the factor.

In order to investigate whether more of the variance could be explained, two-factor and three-factor models were generated using both FML and PFA. The resulting three-factor FML model is shown in [Figure 4](#).

For Factor 1, all variables, except the two types of critical thinking identified in physics, V3 and V4, have quite high, positive loadings. In this factor model, both V3 and V4 are separated out onto individual factors, Factor 2 and Factor 3. This is indicated by the high values of their loadings for these factors, 0.944 and 0.950, respectively, low loadings for the other variables, and low overall uniqueness values for the two variables. The cumulative variance of the model is higher than the proportion variance of the one-factor model, but is explained by the fact that two of the factors are so close to unique, individual variables. A similar loading pattern, with the physics-specific variables V3 and V4 being separated out onto individual factors, was found with PFA. With the two-factor FML model, V3 was separated onto one factor. Models involving more factors cannot be generated because of limitations of the used software, but such models would probably separate out additional variables onto their own factors. Our conclusion is that the factor models therefore have limited use in understanding correlations between the seven variables, however it gives us some indications of patterns of value to discuss and investigate further in future research.

```

Uniquenesses:
  V1  V2  V3  V4  V5  V6  V7
0.654 0.777 0.544 0.679 0.455 0.618 0.518

Loadings:
  Factor1
V1 0.588
V2 0.472
V3 0.675
V4 0.567
V5 0.738
V6 0.618
V7 0.694

                Factor1
SS loadings      2.755
Proportion Var   0.394

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 21.15 on 14 degrees of freedom.
The p-value is 0.0978

```

Figure 3. One-factor FML model, based on the Pearson correlation matrix

```

Uniquenesses:
  V1    V2    V3    V4    V5    V6    V7
0.547 0.650 0.005 0.005 0.495 0.554 0.496

Loadings:
  Factor1 Factor2 Factor3
V1 0.638   0.214
V2 0.583
V3 0.211   0.944   0.244
V4 0.175   0.251   0.950
V5 0.531   0.412   0.231
V6 0.558   0.129   0.343
V7 0.600   0.350   0.144

SS loadings      Factor1 Factor2 Factor3
Proportion var   0.254   0.187   0.166
Cumulative var   0.254   0.441   0.607

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 0.65 on 3 degrees of freedom.
The p-value is 0.884

```

Figure 4. Factor analysis maximum likelihood (FML) with three factors

Concluding discussion

Critical thinking in syllabi and test questions

What we find in this study when we analyse the syllabi and test questions are seven different ways in which critical thinking is operationalized and tested, dimensions we have labelled: (1) *critical thinking in the science of physics*, (2) *critical thinking regarding physics in society*, (3) *critical historical thinking*, (4) *critical thinking about uses of history*, (5) *critical thinking in interpreting and integrating ideas in Swedish*, (6) *critical thinking in examining and evaluating content and form in Swedish*, and (7) *critical thinking in mathematics*. The diversity of formulations and test questions all fall under Facione's (1990) definition of critical thinking, and they contain differences, but also similarities.

Formulations of critical thinking within each school subject focus on aspects found in the corresponding academic subjects, but they also ask students to critically consider aspects a bit outside of the academic tradition. For instance *critical historical thinking* is a traditional way of understanding the past whereas *critical thinking about uses of history* asks students to reason and speculate about the uses of history in ways that academic historians would not normally do in their profession. In physics we find a similar division between critical thinking focusing on the traditional academic subjects and a second type of critical thinking with a focus on society less in line with traditional academic physics.

We also find formulations with great similarities across disciplinary boundaries. Syllabi formulations for critical thinking in Swedish, *critical thinking in examining and evaluating content and form in Swedish*, and history, *critical historical thinking*,

hold great similarities. Both syllabi underscore how it is important to “review and evaluate” information. This focus on reviewing information also resembles formulations in physics asking students to “differentiate facts from values” when dealing with issues linked to physics in society. Formulations in mathematics underscoring the importance for students “to develop their knowledge in order to formulate and solve problems, and also reflect over and evaluate selected strategies, methods, models and results” we find very similar to the viewpoints expressed in *critical thinking in the science of physics* also emphasising the importance to understand models and standards to evaluate processes. Similarities also show relationships within the humanities (history and Swedish) and science (mathematics and physics). However, physics seems to also partly relate to the humanities, or at least social studies, in the dimension we label *critical thinking regarding physics in society*, at least in the formulations in syllabi and in test questions.

Grades in subjects and critical thinking in students’ test results

Our analysis of grades and critical thinking in students’ test scores, on questions testing critical thinking, shows stronger correlations between grades in the subject and critical thinking within the subject than for critical thinking across subjects (see [Table 1](#) and [Figure 1](#)). This means that, for instance, a student with a high grade in physics is likely to think critically *regarding physics in society* (V3), but her critical thinking in physics may not be linked to her *critical thinking in interpreting information in Swedish* (V1). As a conclusion from the correlation analysis, the positive, but low correlations between the seven variables that represent different aspects of critical thinking, both within and across the four different subjects, may be seen as support for the argument that these aspects represent distinct competencies. The generally stronger correlations between the scores on critical thinking and the final grades in the corresponding subjects (with the exception of Swedish), may indicate that critical thinking within a subject is not distinct from general performance or ability in the subject. If you are good at critical thinking within a subject, you tend to get a high grade in that particular subject (and in some cases, also in other subjects).

However, we call for caution in drawing overly farfetched conclusions from this rather limited data set. First of all, the final grades are based on years of interaction and opportunities for assessments between the involved students and their assessing teachers. In contrast, the scores on critical thinking are based on individual, or a few written test items, which were chosen by us as researchers (based on descriptions of what the items were supposed to test), and can be expected to yield higher variability. In addition, the assessment of the test items is based on a less fine-grained-grading scale than the final grades, and assumptions of how to translate them to numerical values.

On the other hand, the validity and reliability of national tests’ items may be considered a bit higher than grades given considering the complexity of giving grades and how hard it is for teachers to grade students with regard to the multiple criteria in the syllabus. In addition, high correlations between scores on critical thinking in a subject and grades in the subject could be seen either as an indication of a causal relationship – critical thinking is a requirement for or leads to general subject knowledge or vice versa – or that critical thinking is identical to general subject knowledge. Further empirical study is needed to clarify these matters.

Complicated relations between formulations, test and test scores

Formulations in the syllabi do not necessarily go hand in hand with the design of test questions and students test scores. Even though the formulations for getting a good grade are similar in the syllabi of Swedish and history, the test results do not co-vary very strongly. We find that students with test scores indicating good *critical historical thinking* (V6) do not necessarily score high on questions testing *critical thinking in examining and evaluating content and form in Swedish* (V2; see [Figure 1](#)). Thus, we find that students able to “review and evaluate” in history fail to do this in Swedish, and students will find that “review and evaluate” means very different things. This needs to be noted to add clarity in teaching and assessments.

Further, formulations of *critical thinking in the science of physics* (V4) and *critical thinking in mathematics* (V5) are quite similar in the syllabi, but the correlations in the test scores are moderate and weaker than correlations between *critical thinking in the science of physics* (V4) and *critical historical thinking* (V6). In this case, correlations between critical thinking in the humanities and the sciences are stronger in the test scores than links between science and mathematics. Perhaps the links between critical thinking in history and physics can, at least partly, be explained by the fact that both dimensions focus on investigating and evaluating the processes of knowledge construction. Another contrast to the syllabi are the fact that correlations are stronger between *critical thinking in the science of physics* (V4) and dimensions of critical thinking in history (V6 and V7) than links between history (V6 and V7) and *critical thinking regarding physics in society* (V3). However, the highest correlation between the results on the critical-thinking categories is found between *critical thinking in mathematics* (V5) and *critical thinking regarding physics in society* (V3); two closely related subjects and with test items that concern applying quantitative thinking to everyday or societal issues (pool fees and energy sources, respectively). Here, the test items are quite similar but the formulations in the syllabi are not as similar as the design of test questions. This means that similar test questions in physics and mathematics are assumed to test different types of critical thinking. The mixed and contrasting relationships between students’ test scores, the design of test questions and formulations in the syllabi highlights how it must be hard for teachers and students to make sense of what critical thinking means within and across subjects. In sum, the same formulation may be transformed into very different test questions in different subjects, what seem to be different skills in the syllabi of different subjects may be similar skills in practice, and similar tests are supposed to test very different critical thinking in different subject. The fact that critical thinking as manifested in test design and students test scores may differ from formulations in syllabi highlights how we need to further investigate in empirical studies what critical thinking may be in theory and practice.

Our statistical analysis complicates the image of critical thinking as a matter of subject-specific skills or general skills even further. On the one hand, it provides arguments for the notion that critical thinking is very diverse on many levels. On the other hand, we may argue that the many links between different dimensions is an indication that, at the end of the day, it is all a matter of a general skill with a number of sub-skills.

Theory, practice and future research

Our results indicate that critical thinking is not clearly a general skill. If this was the case, then there would be high correlations between test questions that have been designed to assess students' critical thinking. This is not the case; instead, good final grades in the subject correlate with good critical thinking scores in the national tests, indicating that critical thinking is an important aspect within each subject. For teachers it may be useful to note that knowledge in the subjects is closely linked to critical thinking in the subjects.

We call for caution regarding what implications can be drawn from our rather limited data set. First, our study focuses on how critical thinking is expressed and operationalised in practice in Swedish compulsory school, rather than on critical thinking as a theoretical construct. This relies on a series of interpretations and translations of what critical thinking is, from the syllabus writers and test designers, to the teachers and us as researchers. Second, it might be that students have been prepared differently for critical thinking across subjects, which may impact on how well they can apply it in the different subjects. Then again, our analysis shows the potential in looking for patterns in readily available test scores across subjects from the authentic school practice. We investigate curricula, syllabi and assessment which are supposed to support and assess critical thinking among all students in Sweden, within and across subjects.

In light of the Ennis (1989) versus McPeck (1990b) debate, we find patterns in our empirical data supporting a perspective that disciplines may hold different dimensions of critical thinking. But even Ennis with a perspective on critical thinking as a general skill acknowledge that disciplines may hold some variation. Ennis and McPeck both state that this is not a clear-cut case. The dimensions of critical thinking that we find in this study need to be further investigated to better understand the complex reality of critical thinking within and across disciplines. The impact of dedicated, domain-general critical-thinking training, and comparison to a baseline of a general critical-thinking test would be interesting to study. In addition, not only do we need to investigate how critical thinking is expressed in tests in more schools across the country. Perhaps even more urgent is to link the complicated relationships we find here to ongoing classroom practices. Following students *in situ* when they think critically within and across subjects can provide us with answers to the questions regarding critical thinking as a subject-specific or general skill and also how this may be furthered by teaching and learning, within and across disciplines.

Acknowledgments

This study was initiated at a network meeting at Stanford University entitled "Critical thinking within and across disciplines in a digital era" and we owe great gratitude to Nobel laureate Carl Wieman, Jo Boaler, Dan Schwartz, Brian Johnsrud, Zephyr Frank, Fred Turner, Sarah Levine and Peter Williamson for sharing their knowledge and perspectives on critical thinking with us.

Funding

This work was supported by the Knut och Alice Wallenbergs Stiftelse;

Notes on contributors

Thomas Nygren is Associate Professor at the department of Education, Uppsala University. His research interests focus on history education, the digital impact on education, critical thinking and human rights education. His previous research, conducted also at Umeå University and Stanford University, has been published in books and journals of history, education, and digital humanities. His current research projects investigate students' news literacy, global citizenship education, and critical thinking across disciplinary boundaries.

Jesper Haglund is a researcher and docent in Physics with Specialization in Physics Education at the Department of Physics and Astronomy, Uppsala University, Sweden. He received his PhD in Science Education from Linköping University, Sweden, in 2012. Haglund's research focuses on the use of metaphors and analogies in science education, and teaching and learning of thermal science at different age levels. In conjunction with his engagement in teaching physics education for pre-service physics teachers, he has recently been involved in research and development on how critical thinking is expressed across different school subjects.

Christopher Robin Samuelsson is a PhD student in Physics Education at the Department of Physics and Astronomy, Uppsala University, Sweden. In his PhD project, he is exploring the teaching and understanding of thermodynamics in laboratory practice across disciplines such as physics and chemistry. In particular, he is studying how students can make use of IR cameras in their investigations of thermal phenomena. Christopher Robin has used quantitative methods for work on undergraduate level and is now using qualitative methods in his PhD project but is interested in the mix of both approaches to education research

Åsa af Geijerstam holds a PhD in linguistics from the Department of Linguistics and Philology, Uppsala University, Sweden and works as an assistant professor in curriculum studies with a focus on mother tongue education at the Department of Education, Uppsala University. af Geijerstam's research focuses on aspects of disciplinary literacy concerning students reading and writing development in differing school subjects. af Geijerstam has been involved in research projects investigating the assessment of reading in international as well as national tests and is currently working with research on content aspects of early school writing.

Johan Prytz is Associate professor (Docent) at Uppsala University, Department of Education. His current research concerns different attempts to reform mathematics education in Sweden in the 20th century. Some of these attempts did involve ideas about formal reasoning and critical thinking. Mathematics was then viewed as an excellent topic to train these abilities, which were considered generic. Other research interests in mathematics education concern the use of language and texts in mathematics education

ORCID

Thomas Nygren  <http://orcid.org/0000-0003-1884-3252>

Jesper Haglund  <http://orcid.org/0000-0003-4997-2938>

Christopher Robin Samuelsson  <http://orcid.org/0000-0003-3070-567X>

Åsa Af Geijerstam  <http://orcid.org/0000-0002-7186-5381>

Johan Prytz  <http://orcid.org/0000-0002-7718-1343>

References

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, 85(2), 275–314.

- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Dordrecht, Netherlands: Springer.
- Broman, A. (2009). *Att göra en demokrat?: Demokratisk socialisation i den svenska gymnasieskolan* (doctoral dissertation). Karlstads universitet, Karlstad.
- Davies, M. (2013). Critical thinking and the disciplines reconsidered. *Higher Education Research & Development*, 32(4), 529–544.
- Eliasson, P., Alvé, F., Yngvéus, C. A., & Rosenlund, D. (2015). Historical consciousness and historical thinking reflected in large-scale assessment in Sweden. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 171–182). New York & London: Routledge.
- Ennis, R. H. (1962). A concept of critical thinking. *Harvard Educational Review*, 32(1), 81–111.
- Ennis, R. H. (1985a). Critical thinking and the curriculum. *National Forum*, 65(1), 28–31.
- Ennis, R. H. (1985b). A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44–48.
- Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, 18(3), 4–10.
- European Commission (2016). A new skills agenda for Europe: Working together to strengthen human capital, employability and competitiveness. Retrieved from <http://ec.europa.eu/social/main.jsp?catId=1223>
- Everitt, B. (2005). *An R and S-plus® companion to multivariate analysis*. New York: Springer.
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations*. Fullerton: US Department of Education.
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., ... Shanahan, C., Project READI. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 219–246.
- Hair, J. F., Anderson, R. E., Babin, B. J., & Black, W. C. (2010). *Multivariate data analysis: A global perspective* (Vol. 7). Upper Saddle River: Pearson.
- Halliday, J. (2000). Critical thinking and the academic vocational divide. *Curriculum Journal*, 11(2), 159–175.
- Herrenkohl, L. R., & Cornelius, L. (2013). Investigating elementary students' scientific and historical argumentation. *Journal of the Learning Sciences*, 3(22), 413–461.
- Hjort, S. (2014). *Kritiskt tänkande i klassrummet: En studie av didaktiska val och manifesterat kritiskt tänkande i samhällskunskaps-och filosofiundervisning* (Doctoral dissertation). Linköping University Electronic Press, Linköping.
- Holmes, N. G., Wieman, C. E., & Bonn, D. A. (2015). Teaching critical thinking. *Proceedings of the National Academy of Sciences*, 112(36), 11199–11204.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Karlsson, K. G. (2014). Historia, historiedidaktik och historiekultur-teori och perspektiv. In K. G. I Karlsson & U. Zander (Eds.), *Historien är närvarande: Historiedidaktik som teori och tillämpning* (pp. 13–89). Lund: Studentlitteratur.
- Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Thousand Oaks: SAGE.
- Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, 67(3), 255–276.
- Lundqvist, E., Säljö, R., & Östman, L. (2013). *Scientific literacy: Teori och praktik*. Malmö: Gleerups.
- McPeck, J. E. (1981). *Critical thinking and education*. New York: St Martin's Press.
- McPeck, J. E. (1990a). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher*, 19(4), 10–12.
- McPeck, J. E. (1990b). *Teaching critical thinking: Dialogue and dialectic*. New York: Routledge.

- Moore, T. J. (2011). Critical thinking and disciplinary thinking: A continuing debate. *Higher Education Research & Development*, 30(3), 261–274.
- Mullis, I., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *PIRLS 2011 assessment framework*. Boston: TIMSS & PIRLS International Study Center Lynch School of Education.
- Nakazawa, M. (2011). R practice: Factor analysis. Retrieved from <http://minato.sip21c.org/swtips/factor-in-R.pdf>
- Namey, E., Guest, G., Thairu, L., & Johnson, L. (2008). Data reduction techniques for large qualitative data sets. In G. Guest & K. M. MacQueen (Eds.), *Handbook for team-based qualitative research* (pp. 137–162). Lanham: Altamira.
- Nygren, T., Sandberg, K., & Vikström, L. (2014). Digitala primärkällor i historieundervisningen: en utmaning för elevs historiska tänkande och historiska empati. *Norddidactica: Journal of Humanities and Social Science Education*, (2), 208–245.
- Nygren, T., Vinterek, M., Thorp, R., & Taylor, M. (2017). Promoting a historiographic Gaze through multiperspectivity in history teaching. In H. Å. Elmersjö, A. Clark, & A. & M. Vinterek (Eds.), *International perspectives on teaching rival histories: Pedagogical responses to contested narratives and the history wars* (pp. 207–228). Basingstoke: Palgrave Macmillan.
- OECD. (2015a). Intervention and research protocol for OECD project on assessing progression in creative and critical thinking skills in education. EDU/CERI/CD(2015)19.
- OECD. (2015b). Assessing progression in creative and critical thinking skills in education. Retrieved from <https://www.oecd.org/edu/ceri/assessingprogressionincreativeandcriticalthinkingskillsineducation.htm>
- Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science*, 328(5977), 459–463.
- Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, 18(1), 16–25.
- Roberts, D. A. (2007). Scientific literacy/science literacy. In S. K. Abell & N. G. Lederman (Eds.), *Handbook on research of science education* (pp. 729–780). Mahwah: Lawrence Erlbaum Associates.
- Rosenlund, D. (2016). *History education as content, methods or orientation?: A study of curriculum prescriptions, teacher-made tasks and student strategies* (Doctoral dissertation). Lunds universitet, Lund.
- Royalty, J. (1995). The generalizability of critical thinking: Paranormal beliefs versus statistical reasoning. *The Journal of Genetic Psychology*, 156(4), 477–488.
- Sá, W. C., Stanovich, K. E., & West, R. F. (1999). The domain specificity and generality of belief bias: Searching for generalizable critical thinking skill. *Journal of Educational Psychology*, 91, 497–510.
- Samuelsson, J., & Wendell, J. (2016). Historical thinking about sources in the context of a standards-based curriculum: A Swedish case. *The Curriculum Journal*, 27(4), 479–499.
- Shanahan, C., Shanahan, T., & Mischia, C. (2011). Analysis of expert readers in three disciplines: History, mathematics, and chemistry. *Journal of Literacy Research*, 43(4), 393–429.
- Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter? *Topics in Language Disorders*, 32(1), 7–18.
- Siegel, H. (2013). *Rationality redeemed?: Further dialogues on an educational ideal*. New York: Routledge.
- Skolverket. (2011). *Curriculum for the compulsory school, preschool class and the recreation centre*. Stockholm: Ordförrådet AB.
- Skolverket. (2013a). *Nationellt ämnesprov i historia, årskurs 9, 2012/2013*. Stockholm: Author.
- Skolverket. (2013b). *Nationellt ämnesprov i svenska, årskurs 9, 2012/2013*. Stockholm: Author.
- Skolverket. (2013c). *Nationellt ämnesprov i fysik, årskurs 9, 2012/2013*. Stockholm: Author.
- Skolverket. (2013d). *Ämnesprov i matematik, årskurs 9, 2012/2013*. Stockholm: Author.
- Smith, G. (2002). Are there domain-specific thinking skills? *Journal of Philosophy of Education*, 36(2), 207–227.
- Solon, T. (2007). Generic critical thinking infusion and course content learning in introductory psychology. *Journal of Instructional Psychology*, 34(2), 95–110.

- Stenhag, S. (2010). Betyget i matematik: Vad ger grundskolans matematikbetyg för information? (Doctoral dissertation). Uppsala universitet, Uppsala.
- Stevens, R., Wineburg, S., Herrenkohl, L. R., & Bell, P. (2005). Comparative understanding of school subjects: Past, present, and future. *Review of Educational Research*, 75(2), 137.
- Tashakkori, A., & Teddlie, C. (Eds.). (2010). *Sage handbook of mixed methods in social & behavioral research*. Thousand Oaks: Sage.
- Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach? *Arts Education Policy Review*, 109(4), 21–32.
- Wineburg, S. S. (1991). Historical problem solving: a study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal Of Educational Psychology*, 83(1), 73. doi:10.1037/0022-0663.83.1.73
- Wineburg, S. S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past*. Philadelphia: Temple University Press.
- Wyndhamn, A. K. (2013). *Tänka fritt, tänka rätt. En studie om värdeöverföring och kritiskt tänkande i gymnasieskolans undervisning* (Doctoral dissertation). Göteborgs universitet, Göteborg.
- Zhao, N. (2009). The minimum sample size in factor analysis. Retrieved from <https://www.encycwiki.org/display/~nzhao/The+Minimum+Sample+Size+in+Factor+Analysis>