2024/3 (19. vol.)
15-36
ERUDITIO-EDUCATIO
Tanulmányok / Studies

Examining the cognitive validity of the intermediate school-leaving exam tasks in History with the thinking-aloud method among secondary school students

Judit Tóth

Abstract
Previous research has shown the difficulty of creating tasks that measure complex thinking. With History, this means that tests do not assess historical thinking (or other high level cognitive processes), but declarative knowledge, reading comprehension or test-solving strategies, making the interpretation of test results problematic. The aim of this pilot study is to explore the extent to which the cognitive requirements stated in the content regulatory documents (e.g., differentiating what has been learned according to causes and consequences) are in line with students' responses (cognitive validity) and whether there are any construct relevant and irrelevant factors that influence task solving. According to hypotheses, the chosen task measures different cognitive processes than stated, and construct relevant, irrelevant factors decrease the validity of the task items as well. The study has been conducted among 12th grade secondary school students (n=37) with the thinking-aloud protocol. The research findings suggest that the task used has raised multiple validity concerns. While most students (83%) successfully completed the task with minimal errors, both construct relevant and irrelevant factors influenced task solving. The need for a comprehensive psychometric development of intermediate tasks in the History school leaving exam is indicated by these results.
Keywords: cognitive validity; thinking-aloud method; school leaving exam; history didactics
Subject-Affiliation in New CEEOL: Social Sciences – Education – School Education
DOI: 10.36007/eruedu.2024.3.015-036

Introduction

In the time of evidence-based policy, in today's globalised educational field, improvements based on accurate measurement-assessment tools are the foundation of educational development and play an important role in quality assurance and accountability, thus determining the effectiveness of the teaching-learning process. In Hungary, the two-tiered school-leaving exam plays a role in providing feedback

on educational outcomes at a given stage of education. The analysis of the tasks of the history school-leaving examination may reveal general methodological-subject pedagogical phenomena that may facilitate the transformation of certain elements of the examination system. Since in the 21st century skills such as problem solving, creativity, digital literacy, and the development of transformative competences (Csapó, 2019) have come to the fore, this study provides insights into the task development culture of the Hungarian history school-leaving system at the intermediate level, pointing out some character traits that define the assessment of historical knowledge as well as influencing secondary education profoundly.

Nature of historial knowledge

"Historical cognition of the sort that enables understanding necessitates far more than memorizing details of what occurred in the past and being able to recall who did what when and where." (VanSledright 2015, 76). In addition to the past, history is also about the knowledge and application of historiographical interpretations (Chapman 2012), i.e. it includes all the knowledge, skills and abilities that facilitate the construction of meaning through the study of history, i.e. the knowledge and adaptation of different patterns and constructions of history through the application of historiographical methods. Although there is a general consensus on certain disciplinary concepts that can enhance historical understanding, there are several competing models of interpreting historical knowledge (Reich 2015; VanSledright 2015), in which, in addition to the divergent models of the German theoretical and Anglo-Saxon empirical traditions (Köstel–Thünnemann–Zülsdorf–Kersting 2019), Spanish research (Gomez & Sáiz 2022), among others, has also come to light. These are all drawn around different understandings of historical thinking and historical literacy (Wineburg 2001; VanSledright 2011, 2012), historical thinking concepts (Seixas–Morton 2013), historical consciousness, adaptive historical literacy, and historical reasoning (Rozendal–Van Boxtel 2023).

Different interpretations have not promoted internationally agreed test preparation standards, so that school-leaving exams in history do not have internationally agreed and widely used test standards, unlike language exams and/or certain science subjects. However, some characteristics of school-leaving exams can be identified: a focus on different levels of cognitive demands, as well as the use of open questions (short answer, long answer, essay type answer) and resource-based (primary, secondary, didactic) tasks/questions. Among the different cognitive models, there are models more widely adapted (e.g. Peter Seixas' historical thinking concepts model), but the almost infinite plurality of models prevents the emergence of historical knowledge as a standard unit. What is peculiar to the history subject is that the expectations of citizens affect the canon of the history subject (Reich 2015).

In the field of history didactics, new models of measurement and assessment have emerged (VanSledright 2011; Resch–Seidenfuß 2017; Rozendal–Van Boxtel

2023; Bertram et al. 2021; Domínguez et al. 2021; Tirado–Olivares et al. 2023), which can serve as theoretical models for multiple interpretations and measurements of historical knowledge. Domínguez et al. (2021) were able to a measure of historical thinking based on the PISA measurement framework for the subject of history in their pilot research. Resch and Seidenfuß (2017) built on the results of earlier German researchers to create a model for the assessment of historical knowledge. Rozendal and Van Boxtel (2023) developed a measurement-assessment model to measure historical reasoning, distinguishing three components: first-order concepts, key interpretive concepts, and epistemological views. The Computational Assessment of Historical Thinking project used artificial intelligence to assess students' open-ended responses (Bertram et al. 2021), including the potential role of Chat-GPT in history teaching (Tirado–Olivares et al. 2023), which has long been of interest in international history didactics research.

Performance tests

Summative assessment (high stakes exams) is a qualification procedure that takes place at the end of the learning and teaching process. Its purpose is to check the level of knowledge achieved by learners and to help education professionals to ascertain whether the objectives set during the learning and teaching process have been achieved. The best-known summative assessment tool is the school-leaving examination, which has a major impact on curriculum content as well as pedagogy, as it influences the curriculum and its methodology through the content that is assessed. Accordingly, if the assessment does not accurately reflect the set of objectives, it limits/distorts the curriculum. School-leaving exams thus function as a kind of hidden curriculum. They emulate the context/circumstances in which students are expected to apply those knowledge and/or skills that are required from them. They are also valuable for educational reforms, as they function as an accountability tool for education policy, as well as a well-designed assessment system: an appropriate tool for managing education and monitoring student learning. Tests, like performance tests, have psychometric properties as measuring instruments, which determine the extent to which the instrument measures reliably, accurately, objectively and in accordance with the test design procedures the constructs measured by the test. For all these reasons, there are predefined, detailed principles of task development. Ignoring/partially considering the principles of item construction will lower certain good-of-fit indicators of the instrument, which are intended to help interpret the test result: these include reliability, validity, objectivity and some researchers also refer to manageability (Smith–Fey 2000). If the psychometric properties of the test are weak, it is not possible to clearly establish exactly what construct the test is designed to measure, which makes it difficult to formulate feedback and, more generally, to assess the effectiveness of both teaching and learning taking place in the educational cycle.

Problems with the assessment of historical thinking: cognitive validity

Cognitive/content validity is a goodness-of-fit indicator of a test, which shows the extent to which the results of the test are interpretable, appropriate, and that the test measures the constructs that the test was intended to measure (Crocker–Algina 2006; Lane 2013; Rodriguez–Haladyna 2013). As a function of this, validity is the most important indicator in task development (Leighton 2017).

Cognitive validity in the assessment of historical knowledge

Previous research has demonstrated (Ferrara 2007; Ferrara et al. 2011) that it is challenging to write tasks that measure complex higher order thinking. "Multiple choice tests tell us only that the correct bubble was blackened, but not what thinking processes led to the choice." (Wineburg 2001, XI) This means that the tests do not/rarely measure historical thinking (or other high-level cognitive processes), but rather declarative knowledge, reading comprehension or test-solving strategies (Ercikan–Seixas 2015; Smith 2017b). Significant differences may therefore exist between the cognitive processes targeted by the task writers and the student actions in real-life task-solving situations. To ensure that the instrument is measuring the constructs it is intended to measure, validity research is needed, and the use of the think-aloud method is common. In their research, Ercikan, Seixas, Lyons-Thomas and Gibson (2015) presented how valid historical assessment tool's multiple-choice questions were able to capture key concepts of evidence, ethical dimension, perspective (historical) and whether students understood the test tasks. The research showed that they understood the task, but different items activated the use of key concepts to different degrees. Reich (2015) also found in his validity research that multiple-choice tests, in addition to the constructs measured, lexical knowledge, reading comprehension, and test-wiseness affected the validity of closed-ended items. Smith (2017a) examined the validity of multiple-choice items of the Historical Thinking Test (HTT) in his research to see how well they measure historical thinking. The research found that, while the test is able to measure historical thinking with some validity, several construct irrelevant factors (e.g., guessing) affect task performance. The National Assessment of Educational Progress (NAEP) standardized history test primarily does not measure historical thinking, but in fact recall and recognition; reading comprehension, or test-solving strategies (Smith 2017b). From Smith's findings, my study has been conducted to see how Hungarian student would solve some task of the intermediate school-leaving exam tasks.

Context of research

The current system of school-leaving examination was introduced in Hungary in 2005.The school-leaving exam is compulsory for all secondary school students (nearly 70,000 students) to be taken in the subject of History as well at least at intermediate level, therefore the methodological-didactic apparatus of the exam determines history classes in secondary education (washback effect). At both levels, the exam consists of a written part (simple short-answer section and essay section) and an oral part, which is modified if changes in content regulations (NCC, exam requirements) happen, both in terms of methodology and content (Katona–Kaposi 2023).

Since 2005, several studies and books have been published on the school-leaving system (Horváth 2010; F. Dárdai–Kaposi 2006, 2008, 2020, 2021; Katona–Kaposi 2023; Kaposi 2015a, 2051b, 2017, 2020, 2023; Makk–Kőfalvi 2007; Csapó 2008; Molnár–Csapó 2019), which describe the system itself and articulate the need for fine-tuning it in general and regarding the subject of history. Among Hungarian researchers, László Kojanitz is of particular importance, who has been working for decades on the complexity of historical knowledge and the assessment of subject-related skills, including the structure of historical thinking and the nature of historical knowledge, and has conducted research on the various tasks of the Hungarian history school-leaving exam (e.g. Kojanitz 2017, 2020, 2021, 2022, 2023).

In the development of the Hungarian history end-of-course examination system, there is a proliferation of practices that call for a change in the culture of writing assignments, on the one hand, and for deficiencies in the culture of writing assignments, on the other. As regards the culture of task development, there is a discrepancy between the written and the assessment curricula, which is reflected in the fact that, apart from the content aspects, cognitive standardisation is not implemented. The didactic role of resources is also questionable, which underlines the presence of pseudo-source-oriented items and justifies the redesign of essay tasks beyond the one-part short-answer task (Kaposi 2023). This is complemented by other psychometric features of the exam (e.g., the formulation of instructions and response options) that suggest the need for fine-tuning the history maturity tasks and their regulatory environment (Kojanitz 2022; Tóth 2022, 2023b). Despite the rich literature and research, we do not have empirical data on Hungarian students' knowledge about history (F. Dárdai–Kaposi 2020). The Educational Authority publishes annual statistics on the school-leaving examination subjects, where results can be found by county, type of school, level, even by student, but we do not know the results of the simple short-answer part of the test, nor of the text (essay) tasks, neither at item nor at task level. In the early 2010s, several large-scale analyses of history subjects were produced (e.g. Csapodi 2014a), but detailed item-by-item analyses would be needed on an annual basis.

Introduction of research

The research[1] was carried out in December 2023 among high school 12th graders in 9 secondary schools (elite high schools, grammar schools, technical colleges) in the capital and rural cities. The sampling method was convenience and quota-based, with an average of 3-4 graduating students from each of the 9 schools having a 4 or 5 as a grade in history. This was necessary because there were two tasks in total included in the study (one open-ended task related to Hungarian history and one closed-ended task related to world history), so I assumed that students with higher grades would be able to answer the open-ended questions as well.

A total of 37 students (15 boys [40.5%], 22 girls [59.5%]) participated in the research, with an average age of 18, 11 (between 17 and 20). Of the 37 students, 11 students (29.7%) were planning to take the advanced level, while 26 students (70.3%) intended to choose the intermediate level in May 2024. Furthermore, 18 students (48.6%) were taking extracurricular classes/other activities in history in their last year.

In the case of the school-leaving examinations in history, there has been no empirical data collected on the historical knowledge and cognitive processes of the students. The aim of this empirical research is to explore the extent to which the task-solving strategies of the intermediate level school-leaving exam align with to the predefined objectives set out in the examination's regulatory documents. Furthermore, how the construct relevant and construct irrelevant processes influence task solving. This study therefore could contextualise the complex range of activities that characterise task development. Construct-relevant processes are those elements that are related to the measurement objectives of the task. Construct-irrelevant processes can be defined as factors that appear in the task solution beyond the construct to be measured, but are not related to the construct to be measured. Most commonly, these are the psychometric characteristics of the task, the test-taking strategies of the test-takers. (Test-taking strategies are procedures that are used to select the correct answer choices during test-taking.) Reading comprehension can be considered as construct irrelevant if, contrary to the intention of the task writer, the task mostly measures only it and therefore does not/hardly measure the construct intended.

The research has been conducted with the students one by one in a quiet room, where they were informed about the research, and also filled in a simple form on their background. After a few warm-up exercises, following the protocol of the think-aloud concurrent interviewing, students were required to solve an one open-ended (Hungarian history, early modern period) and one closed-ended task (world history, modern period) of the intermediate history exam. This paper presents the first results of the closed-ended task, which was included as the 6th task of the May 2006 exam period at intermediate level.

1 Based on the decision of the Ethical Research Committee of the University of Pécs, Humanities and Social Sciences 6/2023 (XII. 4.), this research is granted a research ethics approval.

The tasks were solved on a laptop, projected on a presentation, and the whole assessment took about 20 minutes per student, including the tuning exercises. The coding scheme for the analysis of the audio recordings of the assessment was adapted from the history graduation requirements[2], up to May 2017, and the work of Leighton (2017). Each item was assigned a corresponding valid competency task. These are complemented by cognitive models showing the cognitive structures of the two tasks used in the research. Both the models and the coding scheme help to make the analysis as valid, reliable and standardised as possible.

Research questions:

To what degree do the simple, short tasks in the intermediate school-leaving history exam are in line with the objectives outlined in the regulatory documents?

Which construct-relevant and construct-irrelevant processes influence task solving?

Hypotheses:

The simple, short-answer research tasks of the intermediate school-leaving exam in History is not in line with the objectives outlined in the content regulation documents.

Construct-relevant and construct-irrelevant processes that influence task-solving the most are test-taking strategies, psychometric features of the task (e.g. phrasing), and reading comprehension.

Method:

The think-aloud protocol is a kind of cognitive interview, which is a common method of verbalising cognitive processes and can therefore be used effectively to assess the validity of test scores. Thinking-aloud involves the verbalisation of information processing processes and problem solving, and can therefore be used to reveal whether an item is indeed a measure of problem solving (Lavrakas 2008; Leighton 2017; Padilla–Leighton 2017). Several methodological procedures can be distinguished in the context of thinking aloud, such as concurrent verbal protocol / concurrent interview, restospectice debriefing, and cognitive laboratory (Leighton–Padilla 2017; Leighton 2017; Taylor–Dionne 2000). As this method is time and resource-intensive, these studies under no circumstances can be representative, since these researches involve participants with 15-20 people (Lavrakas 2008; Leighton–Padilla 2017; Leighton 2017; Taylor–Dionne 2000).

The concurrent verbal protocol/concurrent interview is the most reliable method for tracking thinking processes. This is the verbalisation of thinking activity that occurs simultaneously with problem solving. However, these verbal manifestations may not be complete. Only the observed traces of thinking are verbalised, and these verbalisations are also limited by the processing capacity of short-term memory, as information is lost in the process of thinking aloud and reporting it. Furthermore, some information is more likely to be reported than other aspects of

2 (https://ofi.oh.gov.hu/sites/default/files/ofipast/2010/10/tortenelem_kov.pdf)

thinking, typically related to the goal of the research or to the steps of task-solving. In other words, the method reveals structures of thinking that allow the mechanics of problem solving to be tracked. When using concurrent interviewing without other interview combined, there are no clarifying questions formulated by the researcher that interrupt participants' verbalisations (Taylor–Dionne 2000; Leighton 2017).

This paper presents the results using the text summary (Leighton 2017) method, which is a descriptive-exploratory method not requiring coding, and in which themes, patterns, implications and issues identified in the reports can be summarised.

Presentation of the task (measuring instrument)

The task included in the research is the May 2006 task 6 (Table 1[3]) of the simple short-answer tasks of intermediate school-leaving exam, which is about the Second Industrial Revolution. Although this is one of the first modern world history examination tasks ever introduced - these tasks of the first few examination periods differed from those of the other examination periods (e.g. in tems of difficulty, Csapodi 2014b), this task may be ideal for identifying certain thought processes for a number of reasons.

The development of the simple, short-answer tasks of the intermediate history examination are misaligned with the content of the regulatory documents (Kojanitz 2022; Tóth 2022, 2023, 2024), so it is difficult to determine the measurement and assessment objectives of the tasks based on the examination requirements. This task has a clear measurement objective, so the validity of the task can be determined even if, based on task development practice, the tasks are not designed with close attention to the school leaving examination requirements. It also meets the other requirements, as it is a simple(r), authentic task, it can still be described as novel, since although it is an alternative choice task in terms of task-solving activity, it is a task for distinguishing a second-order concept takes place. The task is also authentic in the sense that its psychometric properties are not strong, which is a common feature of Hungarian school-leaving exam tasks in general.

3 The Hungarian version of Table 1 can be found in the Appendix at the end of this study.

The task is related to the second industrial revolution (3 points)

Decide, on the basis of your sources and knowledge, whether the statements refer to the causes or consequences of the second industrial revolution. *Tell me in which column of the table you would put an X. (0,5 points per item.)*

Chronology:[4]

1876. Telefon
1879. Izzólámpa
1882. Elektromos erőmű
1886. Benzinmotoros gépkocsi
1889. Celluloid film
1894. Rádiójelek

	Mezőgazdaság	Ipar	Kereskedelem	Egyéb
1812	27%	30%	20%	23%
1831	28%	35%	15%	22%
1860	16%	36%	48%	–
1895	10%	37%	53%	–

Az éves nemzeti jövedelem megoszlása a gazdasági ágazatok között a XIX. századi Angliában

Statement		Cause	Consequence
a)	The unequal economic development brought about by the second industrial revolution has sharpened the conflict between the great powers.		
b)	Thanks in part to Europe's rich coal and iron ore mines, steel production and transport have grown by leaps and bounds.		
c)	Without the development of credit unions and the fusion of industrial and banking capital, the second industrial revolution would not have unfolded.		
d)	The massive advances in science also played a role in the development of the second industrial revolution.		
e)	Electricity revolutionised public lighting.		
f)	In England, the industrial revolution led to a sharp fall in national income from agriculture.		
Correct answers: a) consequence; b) consequence c) cause; d) cause; e) consequence; f) consequence			

Table 1: The closed-ended task used in the research

4 Translation of the sources: 1876 telephone; 1879 light bulb; 1882 electricity plant; 1886 vehicle with petrol engine; 1889 celluloid film; 1894 radio signals. The columns in bold: agriculture, industry, commerce, other. The title of the table: Distribution of annual national income by economic sector in 19th century England

The teaching of second-order concepts are prominent both in international history didactic discourse and history teaching practice, and they are also included in the National Core Cirruculum (2020) as a primary development goal. These interpretative key concepts are specific analytical skills for the field of history, which can be acquired by students to develop nuanced knowledge of history (Kojanitz 2020). Therefore, this task could be used to examine to what extent this task can measure the existence of knowledge related to the concept. In terms of topic choice, this is a topic that, according to the current content regulation, all final-year students have already learned, but not recently since it is covered in the middle/end of the second year of secondary school history material[5] and is presumably not a favorite topic, meaning students do not have a much broader knowledge of the that than they have learned in school due to their own interests.

Constructs measured by the task:

Using and evaluating sources: gathering information and drawing conclusions from simple statistical tables, charts, graphs, chronologies.

Students can link and distinguish between causes and consequences of the historical factors listed.

Exploring the factors that shape events: organising what has been learned into causes and consequences.

Students can interpret and distinguish statements about the content of statistical tables and chronologies.

Results: Figure 1 shows that the average correctness of the items for most students was 83.3%, i.e. the majority of students were able to solve the task without or with maximum one mistake. This means that for these 37 students, the task was easy. However, it can also be seen that the correct answers for items ‘a’ and ‘b’ are lower. Beyond the correct answers, it is worth looking at the factors that influence the correctness of the answers, which are not revealed by this simple statistic.

5 This topic is covered in grade 10 for general 4-year-secondary schools, in grade 8 for 6-year-secondary schools.

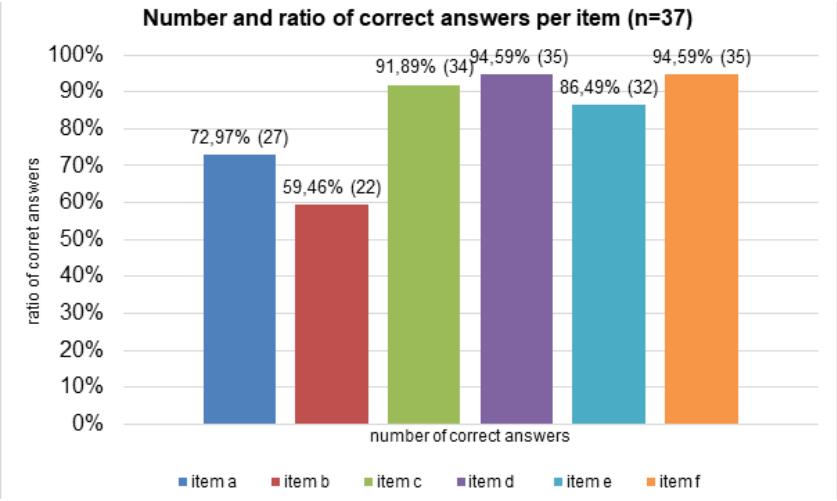

Figure 1: Answers given by students (n=37)

Construct-relevant factors

Since this was a close-ended task, everyone could solve it. However, not all students justified their choice, and e item was the least justified item. According to the verbal reports, students did not/rarely used the sources. Those who reported using them commented on the didactic function of the resources, not understanding what function they serve. *"Well, the sources weren't concrete either, but I don't think they added much to the task, so I think you need a lot more background knowledge than the sources could add..."* (Student 1) Some students were somewhat skeptical about the sources, believing they could solve the tasks on their own. *"So I have to decide whether the given sentence is a cause or a consequence and we have a [sic!] source that might help, but I might try to say first according to my own logic whether it's a cause or a consequence. [...] And then I basically didn't use the source."* (Student 24)

Construct irrelevant factors

Psychometric characteristics are the properties of the test. Such psychometric properties include the wording of the task, in this case the wording of the six statements. As shown in Figure 1, the correctness of items 'a' and 'b' is lower, while the wording of sentence e ("revolutionized") resulted in consequence be given by more than 80% of the students, but they could not justify it. Furthermore, some of the students considered some statements (statements 'a' and 'b' in particular) to be both a cause and a consequence, depending on which clause is considered. *"Well, in this sentence there is a cause, that unequal economic development is*

the cause. The whole sentence itself is probably a consequence, because it is already if there has been an increase or a sharpening of the divide. So I would say it is a consequence." (Student 24, item) *"Well, the thing about this is that part to Europe's rich coal and iron ore mines, steel production and transport have grown by leaps and bounds. This, if I look at it, could be both actually, but a consequence could be that it fits better."* (Student 35, item b)

Another characteristic is that some of the students take the statements as "evidence", as they cannot find anything to be justified. *"Well, this is a consequence. It is [...] quite obvious. Well, it's brought about, it's in the text."* (Student 28, item a) One student further noted that he did not find it unusual that the context implies the solution. *"At least that's how I usually think when I see things like that, so that even if I don't know when [...] it happened, how it happened, there's a chance that I can tell from the wording that it's a cause or a consequence."* (Student 26)

A common test-taking strategy was the disregard of sources and exclusion, most often based on the semantic interpretation of sentences. Answer marking based on intuition also appeared infrequently, but was based on cause-consequence ratios. For example, Student 28 marked item 'd' as a cause in order to have an equal ratio of cause to effect of three to three. His answer is correct, but his assumption is not, for the number of causes is only two.

One student reads the source, but does not use it: *"Again the usual interpretation of the task [read the task, read the source, then the questions], then I look at the source and then the questions. For the tables, I always read exactly what they are about in a meaningful way."* (Student 5)

One third of students misinterpreted the task and/or statements, because they thought they had to list both cause and consequences, even though the instruction tells otherwise. When this occurred, students' justifications greatly varied both in terms first and second order concepts they employed.

There was no significant difference between gender, the chosen level taken and performance. Significant differences between students from different schools were experienced.

Discussion

The justifications behind the answers implies that students have a simplistic view of history, in that they do not report the existence of multiple causes and consequences, but rather that the cause occurred earlier than the consequence. Their failure to provide a nuanced explanation is due to a number of interrelated factors. One may be that students do not have a deep background knowledge of the causes and consequences events of the second industrial revolution (at least they do not remember them). Since they (presumably) do not/should not have, it is quite natural and logical that they try to solve the task and the research request based on the information available to them. This is implied by the fact that they often provide the sentence fragments themselves in their justifications. *"C is also a reason, because without the development, the second industrial revolution would not*

have unfolded." (Student 26, item c) Their view on history can be justified by the curricular focus and expectations, as well as the requirements of the school-leaving exam.

It can also be examined that students with a higher knowledge of history among the 37 students (e.g. OKTV[6] participants, or frequent participants in other competitions) provided more nuanced explanations in their justifications. *"We know that the railways were one of the main driving forces of the industrial revolution."* (Student 18, item b) *"The development of industry and capital is always necessary for some kind of economic change."* (Student 18, item c)

It can be argued that due to the didactic function of the sources included in the exam, students' test-solving strategies in the simple, short-answer task section is ''conditioned'' to solve the task without the source, since it is common to find pseudo-source-based items. These are items in which the source is supposed to be used, but in they require lexical knowledge instead. According to the task instructions, including items that require students to solve them by using their declarative knowledge is a common feature of both exam levels as it is included in the instructions, but poses challenges when the ratio of these items, pseudo source-based items are the most prominent while those items encapsulating higher order thinking are scarce.

Figure 1 shows that items 'a' and 'b' have a lower correctness rate, which has psychometric reasons. The correct answers for items 'c' to 'f' could be inferred semantically by the students, as these contained suggestive words (e.g. ''caused by'').The correct answers for items 'a' and 'b' may also be lower because the words ''brought about'', ''thanks to'', also imply causality, but this was more difficult for the students to interpret. Statement 'e' was the only statement students could rarely justify. This may be because the reasoning was considered too evidential. Indeed, these sentences are mostly explanatory in structure, the different clauses imply to some extent a cause and an effect relationship, and the meaning does not need to be created by the students.

A related problem of justification formulation is also indicated by the fact that in this task a type of subordinate clause is used, where n several cases the first clause expresses the cause and the second the effect or consequence, from which a logical relationship can be inferred. Also, the use of statements as evidence may indicate the formulation of statements, since the task is essentially based on reading comprehension. Because of that, students not only justify their answers by related historical knowledge, but also give a grammatical relations, latter which is more frequent especially among those who seem have to share lover knowledge on the topic.

In this task, the ability to understand the text proved to be the primary measurement construct in which the sources function as distractors, so it is highly implied that for those students whose reading comprehension was lower or who insist-

6 OKTV stands for National Secondary School Academic Competition, the most prestigious competition among secondary school students (ages 15-18) in Hungary which takes place in different subjects, including History.

ed on using the sources in any way had a lower average score. Presumably the students were used to solving the problems in the history exam using the given sources, so one student tried to apply this to all statements and therefore made a mistake. The source for item 'f' could have been used as a help, but only a few students used them. *"I would look at this from the table, and you can see from the figures there that, therefore, in 1812 it still accounted for 27% of the annual national income. But by 1895 it had dropped to 10%. And this is already a consequence of the industrial revolution, so I would list this as a consequence."* (Student 7, item f)

These factors are closely interrelated in the case of this task, since the main psychometric problem of the task is the wording of the statements. In this case, this may be closely linked to the psychometric characteristics of the task, but also implies problems with reading comprehension, which negatively correlates with achievement in other subjects (PISA, 2022[7]). Students solve this task correctly because of its structure and wording, rather than because of prior knowledge of the task topic. This also includes other activities related to the wording of the item (e.g. the student inferring the correct answer from the wording of the item). The majority of students did not use sources to solve the task, despite the fact that the instructions included them. Presumably, if they find the sources to be irrelevant/unhelpful, they skip and solve the task on their own, which was reported by the students as well.

Students solved the task correctly in high proportions, but the factors that emerged during the thinking aloud, both construct relevant and construct irrelevant, imply general psychometric concerns. The results so far suggest that the task is poorly aligned to measure different domains of historical knowledge, mainly due to inadequate wording of the task. In these sentences the logic of order of clauses, first causes and then consequence can be seen, while Hungarian allows for other sentence structures, where these two are reserved in order, making it more difficult for students to find the correct answers merely based on the semantic nature of statements.

This research, although the psychometric characteristics of the task are likely to be weaker than average, supports Wineburg's (2001) idea that we cannot know the exact thought processes behind students' correct (and incorrect) answers. The results of the research therefore raise the question of what kind of history knowledge Hungarian students have and to what extent they are able to apply the skills that the two-tiered history examination tasks claim as a measurement construct. For this reason, the research can be extended and continued in numerous other directions. It raises the question of the nature of the knowledge that Hungarian students have of history. This knowledge is presumably different from what is expected by the history examination syllabus. Regardless of this, the students achieve average scores of 3.5 and 4.4, which may imply that the content-methodological culture surrounding the school-leaving exam in History is known in everyday practice.

7 (https://www.youtube.com/watch?v=b3OWbEQS2A4&ab_channel=MTA1825)

Despite similar grades the history knowledge of these 37 individuals varies greatly, they are not comparable on the same scale. Presumably it is the teacher that plays the most important role in the knowledge students acquire, as several previous research (e.g. Hartmann–Hasselhorn 2008) has indicated.

From a psychometric point, the task is characteristic of construct underrepresentation, because the task does not measure all the constructs identified in the task. Furthermore, irrelevant construct variance (measured construct that are not supposed to be measured) can be also found if we accept that text comprehension was not the primary construct measured in the task design. Students' test-taking strategies, strategies for source use and source criticism, and routines acquired during task-solving also influence the results which require further research. One can look at what other content regulators (e.g. textbooks, framework curricula), how the material was covered in class, how much time was spent on the topic in general. As the nature of historical knowledge does not allow for drawing clear conclusions with close-ended items, VanSledright (2015) suggests the introduction of weighted multiple-choice items, which are considered to be applicable in the history subject, to weight this nuanced place. Finding a possible solution to create valid closed-ended history tasks.

Limitations

Based on the sampling, quota selection by grades is not reliable. Although the survey included students who had a 4-5 in History, some students significantly over- or underperformed compared to the 37 students. Thus, in the future this sampling method can be omitted. Regarding the method, data processing from the thinking-aloud is time-consuming and costly, and not everyone is capable of thinking aloud. Research facilitates verbalisation if the researcher does not carry out the measurement themselves, thus reducing anxiety (Leighton 2017), but this was not possible in this research. Besides, effective use of the thinking aloud method requires problem solving (Leighton 2017; Lavrakas 2008), which is unlikely to occur in a compulsory, high-stakes exam task, as it would have not been democratic. Another shortcoming of the history examination is the lack of clear assessment objectives that can be broken down into tasks/items. Two objectives to this task have been assigned, but we do not know exactly what the task's author had in mind. Thus, given the hypothesized measurement objectives, the task is not valid, but the task writer may have achieved his objective with the task.

The task was chosen based on the topic, the nature of the task (closed), the constructs measured (second-order concepts, didactic function of sources), because it has international relevance. However, with these parameters, the choice was made for a task with psychometric characteristics weaker than an average intermediate level simple short-answer task, which to some extent distorted the complexity of the reports and the nuanced possibility of demonstrating validity. Furthermore, the selected task was not based on the 2020 content requirements, as they are earlier, but this could not have significantly affected the degree of ver-

balisation, as the topic also occurs in todays' content regulatory context. However, it is inevitable to repeat the research in the future with a restrospective debriefing. This was not done in this research because the aim was to obtain the most "natural" (objective) answers possible. By asking additional questions afterwards each verbalisation, it is possible to gain a deeper insight into certain aspects of students' thinking, such as how and why they solved items the certain way, which reveals much more about their actual knowledge of history than just typing in the correct answer. In addition, in the general data collection, it would have been worth asking when they last studied the topic of the task, as this would certainly have influenced the complexity of the justifications.

Conclusion

The data collected in this research provide important and valuable results in understanding the discrepancies between the requirements of testing practices (what is learnt for the exam), and what exactly is to be historically literate. Although the results by no means representative, they imply the issues surfacing on system-level history education in Hungary. Analyses so far have shown that the simple, short-answer task of the intermediate history examination is not well aligned to the objectives formulated in the regulatory documents of the examination. These objectives are related to the ability to distinguish causes and consequences and the ability to use/analyze the sources included in the task. Therefore, it is implied that task I not adequate for valid measurement of these objectives. Among the construct-relevant and construct-irrelevant processes, test-solving strategies, psychometric features of the task (e.g. wording), and reading comprehension are the most significant.

Although the research has many limitations, it has provided new information about learners' thinking and knowledge. The analysis of the data is still in its early stages and needs to be completed and coded. Therefore, As an extension of the research, the data will be coded (pattern coding, theme coding, cognitive coding), presented in a coding scheme and analysed in a qualitative content analysis software (MAXQDA). The method can be extended and new foci can be formulated (e.g. correlation between students' historical knowledge and the test-solving strategies).

Overall, the results so far show the psychometric shortcomings of the selected task. Aspects to be considered for task development: objectives, theoretical basis, which national/international model is adapted for knowledge assessment, what are the areas to be measured, proportion, frequency, weighting of knowledge and skill-based items. The data collected in this research provide important and valuable results in understanding the discrepancies between the requirements of testing practices (what is learnt for the exam), and what exactly is to be historically literate. Although the results by no means representative, they imply the issues surfacing on system-level history education in Hungary.

Taking all this into account, the research is of fundamental importance for increasing the effectiveness of history education in Hungary and for further refining

the examination levels. The results of the research can also be used in the development of exam requirements, textbook writing and teaching practices, and help to put the task development culture of the Hungarian history graduation examination into a broader context. In addition to informing history teachers, policy-makers and other stakeholders, the results can also be used to inform the development of content regulators and the school-leaving exam. Professionally constructed, tested, valid and reliable tests are inevitable for the interpretation of test results and thus for the renewal of the content and methodology of secondary school history teaching. However, finding systemic solutions to problems requires the integrated cooperation of history teachers, task writers, researchers in history didactics and other decision-makers.

Acknowledgements

This study was supported by the "ÚNKP-23-3-II-PTE-2017 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund."

I would like to thank Dr. Mark Smith (Stanford University) for his valuable advice and for contributing to the success of this research. I would also like to thank my supervisor, Dr. József Kaposi (Péter Pázmány Catholic University), Prof. Dr. Susanne Popp (University of Augsburg), Dr. Isidora Sáez- Rosenkranz (University of Barcelona) and Dr. Maja Majkić for their insights.

Appendix: The Hungarian version of Table 1

A feladat a második ipari forradalommal kapcsolatos. (3 pont)
Döntse el a források és ismeretei alapján, hogy a megállapítások a második ipari forradalom okaira vagy következményeire vonatkoznak! Mondja el, a táblázat melyik oszlopába tenne X jelet! (Elemenként 0,5 pont.)

Kronológia:

1876. Telefon
1879. Izzólámpa
1882. Elektromos erőmű
1886. Benzinmotoros gépkocsi
1889. Celluloid film
1894. Rádiójelek

	Mezőgazdaság	Ipar	Kereskedelem	Egyéb
1812	27%	30%	20%	23%
1831	28%	35%	15%	22%
1860	16%	36%	48%	–
1895	10%	37%	53%	–

Az éves nemzeti jövedelem megoszlása a gazdasági ágazatok között a XIX. századi Angliában

Megállapítás		Ok	Következmény
a)	A második ipari forradalom által előidézett egyenlőtlen gazdasági fejlődés kiélezte a nagyhatalmak közötti ellentéteket.		
b)	Európa gazdag szén- és vasércbányáinak is köszönhetően az acélgyártás és a közlekedés ugrásszerűen fejlődött.		
c)	A hitelszervezetek fejlődése, az ipari és a banktőke összeolvadása létrejötte nélkül nem bontakozott volna ki a második ipari forradalom.		
d)	A tudomány nagyarányú fejlődése is szerepet játszott a második ipari forradalom kibontakozásában.		
e)	Az elektromosság forradalmasította a közvilágítást.		
f)	Angliában az ipari forradalom hatására a mezőgazdaságból származó nemzeti jövedelem jelentősen visszaesett.		

Forrás: Oktatási Hivatal – Érettségi feladatsorok, 2006 május (feladatlap): https://dload-oktatas.educatio.hu/erettsegi/feladatok2006tavasz/kozep/k_tort_06maj_fl.pdf (2023. 11. 08. 14:51) megoldókulcs: https://dload-oktatas.educatio.hu/erettsegi/feladatok-2006tavasz/kozep/k_tort_06maj_ut.pdf (2023. 11. 08. 14:52)
Helyes válaszok: a) Következmény; b) Következmény; c) Ok; d) Ok; e) Következmény; f) Következmény

Bibliography

Bertram, C. – Weiss, Z. – Zachrich, L. – Ziai, R. (2021): *Artificial intelligence in history education. Linguistic content and complexity analyses of student writings in the CAHisT project (Computational assessment of historical thinking)*. Computers and Education: Artificial Intelligence.

Bruner, Jerome (2004): *Az oktatás kultúrája.* Budapest: Gondolat Kiadó.

Crocker, Linda – Algina, James (2008): *Introduction to Classical and Modern Test Theory*. Cengage Learning.

Chapman, Arthur (2012): Developing an Understanding of Historical Thinking Through Online Interaction with Academic Historians: Three Case Studies. In: *International Journal of Research on History Didactics, History Education and History Culture*, 32, 21-40. Frankfurt: Wochenschau Verlag.

Csapodi, Zoltán (2014a): *A nagy létszámú vizsgatárgyakra vonatkozó felmérés elemzése. (Történelem)*. Budapest: Nemzeti Fejlesztési Ügynökség. 564-612, 616-630, 703-715.

Csapodi, Zoltán (2014b): *Érettségi vizsgatárgyak elemzése: 2009-2012 tavaszi vizsgaidőszakok. Történelem*. Budapest: Nemzeti Fejlesztési Ügynökség.

Csapó, Benő (2019): A jövő elvárásai és a tudás minősége. In*: Az oktatás átalakulása a tudástársadalom és a mesterséges intelligencia korában: XXI. Országos Közoktatási Szakértői Konferencia.* Debrecen: Suliszerviz Oktatási és Szakértői Iroda.

Csapó, Benő (2008): A közoktatás második szakasza és az érettségi vizsga. In: *Zöld könyv a magyar közoktatás megújításáért.* 71-93. Budapest: Ecostat Gazdaságelemző és Informatikai Intézet.

Domínguez-Castillo, J. – Arias-Ferrer, L. – Sánchez-Ibáñez, R. – Egea-Vivancos, A. – García-Crespo, F. J. – Miralles-Martínez, P. (2021): A competence-based test to assess historical thinking in secondary education: Design, application, and validation. *Historical Encounters*, 8(1), 30-45.

Ercikan, Kadriye – Seixas, Peter (eds.) (2015): *New Directions in Assessing Historical Thinking*. New York: Routledge. Introduction.

F. Dárdai, Ágnes – Kaposi, József (2008): A kezdet vége, avagy a történelemérettségi vizsga fejlesztésének további lépései. In: Bánkuti Zsuzsa – Lukács Judit (szerk.): *Tanulmányok az érettségiről. Hatásvizsgálat. Tantárgyi vizsgák értékelése. Feladatfejlesztés*. Budapest: Oktatáskutató és Fejlesztő Intézet, 169-186.

F. Dárdai, Ágnes–Kaposi, József (2020): Változó történelemtanítás Magyarországon (1990–2020) Trendek, mozaikok, mintázatok. In: F. Dárdai–Kaposi–Katona (szerk.) *A történelemtanítás a történelemtanításért. Válogatás a Történelemtanítás online folyóirat írásaiból (2010–2020)*. Budapest: Magyar Történelmi Társulat, 15–59.

F. Dárdai, Á. & Kaposi, J. (2021): Changing History Teaching in Hungary (1990- 2020). In: *International Journal of Research on History Didactics, History Education and History Culture (JHEC) Yearbook of the International Society for History Didactics (ISHD)*: History Education 30 Years after the Cold War / History Education in Africa, 79-97.

F. Dárdai, Á. & Kaposi, J. (2022): Trends in changing history teaching in Hungary (1990–2020). In: *Hungarian Educational Research Journal*, 12 (2), 164-192.

F. Dárdai, Ágnes – Kaposi, József (2006): Merre tovább történelemérettségi? Javaslatok az új történelemérettségi továbbfejlesztésére. In: *Új Pedagógiai Szemle.* 56. évf. 11. sz., 21-35.

Ferrara, Steve (2007): Toward a Psíchology of Large-Scale Educational Achievement Testing: Some Features and Capabilities. In: *Educational Measurement: Issues and Practice. Introduction.*

Ferrara, Steve & Svetina, Dubravka & Skucha, Sylvia & Davidson, Anne H. (2011): Test Development with Performance Standards and Achievement Growth in Mind. In: *Educational Measurement: Issues and Practice*, Vol. 30, No. 4, 3-15.

Gómez Carraso, Cosme Jesús - Sáiz Serrano, Jorge (2022): The origin and development of research into historical thinking. In: Gómez Carraso – Cosme Jesús (ed): *Re-imagining the Teaching of European History. Promoting Civic Education and Historical Consciousnes.* Routlegde, 25-41.

Harlen, Wynne (2016): Assessment and the Curriculum. In: Wyse, Dominic - Hayward, Louise - Pandya, Jessica (eds.): The SAGE handbook of Curriculum, Pedagogy and Assessment. Sage, 963-709.

Hartmann, Ulrike - Hasselhorn, Marcus (2008): Historical perspective taking. A standardized measure for an aspect of students' historical thinking. In: *Learning and Individual Differences*, 18, 264-270.

Horváth, Zsuzsanna (2010): Az érettségi kérdőjelekkel. Mérleg 2006–2010, *Educatio*, 2010, 54-64.

Kaposi, József (2015a): Válogatott tanulmányok I.: történelem, érettségi, megújítás. Budapest: Szaktudás.

Kaposi, József (2015b): Válogatott tanulmányok II.: tanterv, történelem, módszertan. Budapest: Szaktudás.

Kaposi, József (2017): A történelmi gondolkodás és a képességfejlesztő feladatok. In: *Történelemtanítás: Online történelemdidaktikai folyóirat* (LII.) Új folyam VIII. 1-2.

Kaposi, József (2020): *Közelítések a történelemtanítás elméletéhez és gyakorlatához. Oktatási segédlet.* Budapest: Pázmány Péter Katolikus Egyetem.

Kaposi, József (2023): Az esszéfeladatok problematikája a történelemérettségi vizsgán. In: *Rendszerváltó Szemle (8)*1-2, 46-53.

Kaposi, József – Katona, András (2023): Forrásközlések történelemtanításunk közelmúltjából VI. A történelemérettségi 2024-től érvényes lexikai anyaga – a korábbiakkal összehasonlítva. In: *Történelemtanítás: Online történelemdidaktikai folyóirat* (LVIII.) Új folyam XIV. 4.

Kojanitz, László (2017): A történelmi gondolkodás fejlesztése az Újgenerációs tankönyvekkel. In: *Történelemtanítás: Online történelemdidaktikai folyóirat* (LII.) Új folyam VIII. 3–4.

Kojanitz, László (2020): Az értelmező kulcsfogalmak tanítása és tanulása I. Okok. In: *Történelemtanítás: Online történelemdidaktikai folyóirat* (LV.) Új folyam XI. 3-4.

Kojanitz, László (2021): A történelmi gondolkodás fejlesztése: Válogatott tanulmányok. In: *Belvedere Meridionale*, Szeged.

Kojanitz, László (2022): Érettségi feladatok és forráshasználat. Az érettségi feladatok elemzése az elsődleges források használata szempontjából. In: *Történelemtanítás: Online történelemdidaktikai folyóirat*. (LVII.) Új folyam XIII. 3-4.

Kojanitz, László (2023): A történelmi fogalmak tanításának nehézségei. In: *Történelemtanítás: Online történelemdidaktikai folyóirat* (LVIII.) Új folyam XIV. 3.

Köster, Manuel – Thünemann, Holger – Zülsdorf-Kersting, Meik (2019): International History Education Research: Common Threads, Research Traditions and National Specifics. In: Köster, Manuel – Thünemann, Holger – Zülsdorf-Kersting, Meik (eds): *Researching*

History Education International Perspectives and Disciplinary Traditions. Wochenschau Geschichte, 5-16.

Lane, Susanne (2013): Performance Assessment in Education. In: Geisinger, Kurt F. – Bracken, Bruce A. – Carlson, Janet F. – Hansen, Jo-Ida C. – Kuncel, Nathan R. – Reise, Steven P. – Rodriguez, Michael C. (eds): *APA Handbook of Testing and Assessment in Psychology, Vol. 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology.* American Psychological Association, 329-339.

Lavrakas, Paul J. (ed.)(2008): *Encyclopedia of survey research methods.* (Vols. 1-0). Sage Publications, Inc.

Leighton, Jacqualine, P. (2017): *Using Thinking-Aloud Interviews and Cognitive Labs in Eduational Research.* Oxford University Press, NY.

Leighton, Jacqueline P. & Padilla, José-Luis (2017): Cognitive Interviewing and Think Aloud Methods. In: Zumbo D. Bruno, Hubley M. Anita (eds.): *Understanding and Investigating Response Processes in Validation Research.* Social Indicators Research Series, Vol. 69, pp. 211-228.

Makk, Ferenc–Kőfalvi, Tamás (2007): *Forrástani ismeretek történelemből: Segédtankönyv a történelem forrásközpontú tanításához.* Budapest: Nemzeti Tankönyvkiadó.

Molnár, Gyöngyvér–Csapó, Benő (2019): A felsőoktatási tanulmányi alkalmasság értékelésére kidolgozott rendszer a Szegedi Tudományegyetemen: elméleti keretek és mérési eredmények. In: *Educatio*, 28, 4, 705-717.

Resch, Mario – Seidenfuß, Manfred (2017): A taxonomic analysis of learning tasks in history lessons: Theoretical foundations and empirical testing. In: *International Journal of Research on History Didactics, History Education and History Culture (JHEC) Yearbook of the International Society for History Didactics (ISHD):* History Teacher Education Facing the Challenges of Professional Development in the 21st Century, 235-252.

Reich, Gabriel, A. (2015): Measuring Up? Multiple-Choice Questions. In: Ercikan, Kadriye – Seixas, Peter (eds.)(2015): *New Directions in Assessing Historical Thinking*. New York: Routledge. 221-232.

Rizvi, Fazal (2009): Globalisation and Policy Research in Education. In: Ryan, Katherine E. - Cousins, J. Bradley (eds.): *The SAGE international handbook of educational evaluation.* SAGE Publications, Inc.

Rodriguez, C. Michael – Haladyna, M. Thomas (2013): Objective Testing of Eucational Achievement. In: Geisinger, Kurt F. – Bracken, Bruce A. – Carlson, Janet F. – Hansen, Jo-Ida C. – Kuncel, Nathan R. - Reise, Steven P. - Rodriguez, Michael C. (eds.): *APA Handbook of Testing and Assessment in Psychology, Vol. 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology.* American Psychological Association, 305- 314.

Rozendal, U. D. – Van Boxtel, C. (2023): Illuminating historical causal reasoning: Designing a theory-informed cognition model for assessment purposes. In: *Historical Encounters*, 10(1), 60-75.

Rupp, André A. – Leighton, Jacqueline P. (2017): *The Handbook of Cognition and Assessment. Frameworks, Methodologies, and Applications.* Introduction. Wiley Blackwell.

Seixas, Peter – Morton Tom (2012): *The Big Six Historical Thinking Concepts.* Toronto: Nelson.

Smith, Mark, D. (2017a): Cognitive Validity: Can Multiple Choice Items Tap Historical Thinking Processes? In: *American Educational Research Journal*. 54 (6), 1256-1287.

Smith, Mark (2017b): *New Multiple-Choice Measures of Historical Thinking: An Investigation of Cognitive Validity,* Theory & Research in Social Education.

Smith, M. L. – Fey, P. (2000): Validity and Accountability in High-Stakes Testing. In: *Journal of Teacher Education*, 51 (5), 334-344.

Taylor, Lynn, K. – Dionne, Jean-Paul (2000): Accesing Problem-Solving Strategy Knowledge: The Complementary Use of Concurent Verbal Protocols and Retrospective Debriefing. In: *Journal of Educational Psychology*, 92 (3), 413-425.

Tirado-Olivares, S. – Navío-Inglés, M. – O'Connor-Jiménez, P. – Cózar-Gutiérrez, R. (2023): From Human to Machine: Investigating the Effectiveness of the Conversational AI ChatGPT in Historical Thinking. In: *Education Science*, 13, 803.

Tóth, Judit (2022): A 21. századi készségek-képességek mérése a középszintű történelemérettségi komplex tesztfeladatainak tükrében. [Measuring 21st century skills and competences in the light of the complex source analsation tasks of the Intermediate school leaving exam in History] In: Steklács, János–Molnár-Kovács Zsófia (szerk.): *21. századi képességek, írásbeliség, esélyegyenlőség. Absztraktkötet. XXII. Országos Neveléstudományi Konferencia.* Pécs: MTA Pedagógiai Tudományos Bizottság – PTE BTK Neveléstudományi Intézet.

Tóth, Judit (2023): A középszintű történelem írásbeli érettségi egyszerű, rövid választ igénylő vizsgarész feladattípusainak mintázatai. [Task type patterns of the simple, short-answer part of the written intermediate school-leaving exam in History] Szeged, 2023. April 20-22., XIX. Pedagógiai Értékelési Konferencia. (Lecture)

Tóth, Judit (2024): Examining cognitive processes among secondary school students when solving tasks of the intermediate school-leaving exam in History: A cognitive validity study. Pécs, 2024. April 5-6. XII. Interdiszciplináris Doktorandusz Konferencia. (Lecture)

VanSledright, Bruce (2015): Assessing for Learning in the History Classroom. In: Ercikan, Kadriye – Seixas, Peter (eds.)(2015): *New Directions in Assessing Historical Thinking*. Routledge, New York, 75-88.

VanSledright, A. Bruce (2012): *The Challenge pf Rethinking History Education. On Practices, Theories, and Policy*. Routledge.

VanSledright, A. Bruce (2011): *Assessing Historical Thinking & Understanding. Innovative Design for New Standards*. Routledge.

Vajda, Barnabás: Teaching history in Slovakia. In: Fink, Nadine – Furrer, Markus – Gautschi, Peter (eds.): *Why History Education?* Frankfurt am Main: Wochenschau Verlag, 2023, 138-152.

Wineburg, Sam (2001): *Historical Thinking and Other Unnatural Acts. Charting the Future of Teaching the Past.* Philadelphia: Temple University Press.