

Minne, morot, eller helt enkelt i vägen?  
Visualisering av kunskap i digitala läro-  
medel.

Token, carrot, or just in the way?  
The challenge of visualizing acquired  
knowledge in the era of digital learning  
and gamification.

Eva-Maria Ternblad

supervisor:  
Agneta Gulz

KOGM20

2017-09-19

# Token, carrot or just in the way?

## The challenge of visualizing acquired knowledge in the era of digital learning and gamification.

Eva-Maria Ternblad  
[eternblad@gmail.com](mailto:eternblad@gmail.com)

*Metacognition is necessary for learning. Without knowing what you know, what you don't know and what you are about to learn, seeking new knowledge becomes both hard and inefficient. At the same time, keeping old facts (as well as old skills) in mind at a meta-level when striving for new insights is not always an easy task. Consequently, external visible cues and representations are essential, reminding us of not only of what we have learnt, but also of where we are in the process and where are heading.*

*At present, improving metacognition is on top of the educational agenda for many schools and universities. However, in the current new era of digital learning, the impact from digital tools on metacognitive processes - especially from educational games and apps - is rarely discussed. For instance, in contrast to traditional learning material (books etc.), most applications let the students solve tasks without saving any traces of the solutions, resulting in a minimum of durable external representations and memories.*

*To address the need for metacognitive support in instructional software, the present study examines the effect of a **visualization tool** on knowledge monitoring and self-regulation. The tool, that consists of a diary where tokens with knowledge related content are received as proofs of achievement for solved tasks, was designed specifically for the study at hand and implemented in an educational game (*Guardians of History*). It was tested with a between-subjects design, where 117 Swedish students in grade 5 and 6 played two varieties of the game – one with and one without the diary. Although no significant positive impact of the tool was found, the study reveals several important and interesting findings regarding the challenges of visualizing acquired knowledge in instructional software.*

## 1 Introduction

The digitalization of learning environments and the production of serious games and instructional software is one of the major trends within modern education. Stretching from E-portfolios and tools for organizing and sharing school-related information to fun and entertaining learning games, the areas of application are both numerous and diverse. The multifunctional and multimodal properties of digital technology not only make it possible to individualize instructions and tasks, it can also facilitate the distribution of material and simplify correction and administration. However, since computers and

technical communication tools have different physical properties than traditional paper-based artefacts they also afford different ways of manipulation and perception. The constancy of physical objects in comparison with the flexibility in digital ones hereby also leads to different constraints and affordances in actual learning situations.

Two interesting cognitive aspects with impact on learning efficacy is the capacity of monitoring prior knowledge and the use of effective self-regulating learning strategies. Substantial research implicates that these metacognitive abilities – knowing what you know and studying accordingly – could be facilitated by proper tools and teaching methods. One way of doing this is by visualizing the learning process and making the students' aware of their acquired knowledge in relation to goals and objectives (e.g. see Hattie, 2016; Håkansson, 2011). However, in many educational applications, not at least in games, this functionality is lacking. Instead, the user here typically performs a series of exercises, without saving any substantial traces of them, and without exactly knowing what is up ahead. After finishing the tasks, nothing is saved, and the possibility of reviewing the content or the obtained skills is very limited.

To the author's knowledge, these matters are neither discussed within the field of educational technology, nor are they investigated within experimental settings. Subsequently, the aim of this thesis, is to explore and evaluate the possible effects of visualizing acquired knowledge in instructional software. The software used is *Guardians of History*, an application for elementary students in the 4<sup>th</sup> to 6<sup>th</sup> grade developed by the universities at Lund and Linköping in Sweden. The functionality of saving and presenting knowledge-related information – in the shape of a travel diary with gathered historical souvenirs - has been implemented specifically for the study at hand. The research questions are the following:

- Will a function that saves and presents traces of the user's performed tasks and obtained knowledge catch the user's attention, and will it be used during play?
- Can interaction with this visualized information lead to improved knowledge monitoring and metacognitive processing?
- Could the visualizations also serve as proofs of performance and lead to higher engagement?

The structure of the thesis is as follows: First, a series of passages describing theoretical standpoints concerning learning, cognition and metacognition are presented, to be followed

by a discussion on the design and use of educational technology and its advantages and flaws. The used software, Guardians of History, is briefly described and related to research questions and hypotheses, after which the experimental methodology is reported. The results are presented, quantitatively as well as qualitatively and finally, the project and problem area is discussed and reflected upon. Throughout the entire thesis, a distributed view of cognitive and metacognitive processes is emphasized. Educational tools with specific affordances and properties are thereby evaluated in respect to their capacity of affecting mentalization, memory and the construction of inner concepts and self-beliefs.

### *Theoretical perspectives of cognition, learning and artefacts*

How do we learn? How do activities, experiences and presented facts about the external world turn into knowledge, skills and expertise? And moreover, what instruments and procedures should be used to optimise learning so that teaching and tutoring make students engage in the process and become eager knowledge-seekers as well as skilled practitioners? Even though learning mechanisms are widely studied in a variety of research areas, such as associative learning and conditioning within animal cognition, problem solving strategies and heuristics in the field of artificial intelligence and the use of effective pedagogical tools in the Vygotskian era of developmental psychology, no general or universal answers have been found. In addition, the field of educational psychology is heavily heterogeneous, resting on a diversity of ideologies and practises. Consequently, the different educational traditions do not only differ in their pedagogical recommendations, they also carry out different types of studies, have different levels of analysis and focus on separate outcomes (for an excellent and thorough review on these matters, see Greeno, Collins and Resnick, 1996).

Since this thesis has the ambition to embrace a pragmatic and holistic view of learning by making a synthesis of several paradigms, a quick overview of three of the most important perspectives of educational theories – the behaviourist/empiricist, cognitive/rationalist and the situative/pragmatic-socio-historic view - is given (these concepts correspond to the terminology used by Greeno et al, 1996). The conclusive theoretical lens through which the project is carried out and analysed is then finally described and linked to the current research questions.

According to the *behaviourist/empiricist* approach (stemming from the 1950's), learning could be regarded as the strengthening and/or weakening of associative neural connections by the impact of external stimuli. In other words, the agent memorizes a certain behaviour as pleasant/rewarding or uncomfortable/punishing due to reinforcements (Greeno et al, 1996; Hall, 2002; Shettleworth, 2013). The key to successful learning here lies primarily in instruction, repetition and appropriate responses in relation to performed trials and errors. Using the individual agent as the main unit of analysis and focusing on observable behaviour, behaviourists generally avoid speculating on collaborative, contextual, interactive or mental

aspects or inner states. To optimize knowledge acquisition, this theoretical paradigm recommends individualized instruction and clear goals with adequate scaffolding, assessment of subgoals and immediate formative feedback (Greeno et al, 1996; Ohlsson, 2008; Shute, 2007). No matter how narrow and out-of-date this approach may seem in modern schools, instruction, repetition and feedback are still considered to be important educational elements (Ohlsson, 2008; Hattie & Timperley, 2007). According to the research agenda in this thesis, these elements can also be facilitated, as well as obstructed, through the interactive properties in educational tools.

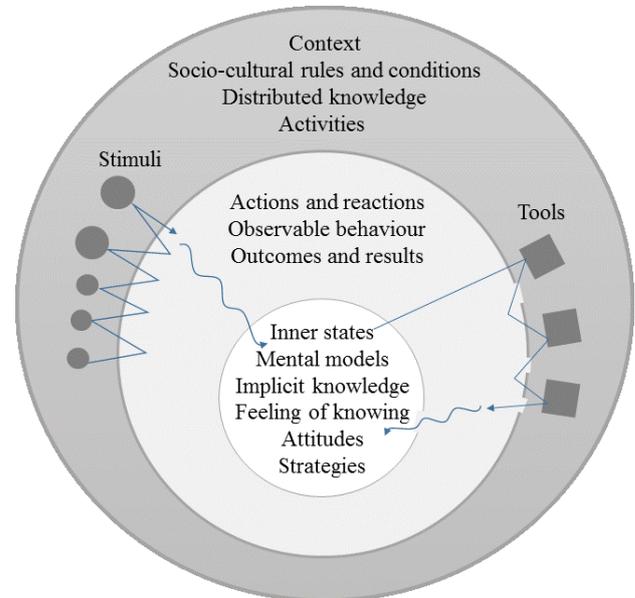
A slightly newer educational approach, mainly growing out of the computational modelling of skill acquisition in the 1980's, is the *cognitivist/rationalist* view, also referred to as a *constructivist* learning theory. According to this perspective, learning does not only consist of plain memorising, but should rather be understood as “a constructive process of conceptual growth” (Greeno et al, 1996, p. 16). Important aspects of learning are in this case the organisation and structure of concepts in the adept's mind and his or her logic and deductive reasoning. The process is assumed to go from practical problem-solving to mentalization of models and from that to transfer and life-long learning. The focus in this paradigm lies thereby in the learning and use of methods and strategies, including self-regulation and meta-cognitive training. In designing learning environments, the perspective advocates interactive settings where the students can construct new knowledge and get attentive towards principles of generality (as opposed to repeating and memorizing specific facts) (Greeno et al, 1996; Gärdenfors, 2010). Focusing on the mental processes and inner structures of symbolic representations of the learning agent, the unit of analysis is here the same as in the behaviourist perspective, while the level is slightly different. The framework has promoted a series of useful investigations concerning meta-cognitive abilities, learning styles and problem-solving methods. However, the possibility of teaching general abilities - at the expense of factual knowledge - has been debated and criticised during the latest decades (Linderoth, 2010; Gärdenfors, 2010). Furthermore, the automaticity of constructing mental models from perceptual input or exploration should not be overestimated, and it is important to bear in mind that human reasoning as well as learning rely on an efficient but parsimonious cognitive economy, where deduction and logic takes time and effort while motor tasks, the use of tools and the search for visual cues are less demanding (de León, 2003; Kirsh, 1996; Kirsh & Maglio, 1992; Linderoth, 2012). Since the study at hand aims to evaluate metacognitive aspects of digital learning, the constructivist paradigm is of great interest. Still, the question of which ingredients a software must have to reach this level of understanding, remains to be answered.

Finally, a third way of studying learning processes is by applying a *socio-historic* or *pragmatic* perspective. This view takes not only the individual into consideration, but widens the analysis to include the socio-cultural context, peers and the usage of tools and teaching material (Greeno et al, 1996). In this

paradigm, knowledge and learning are treated as distributed processes within a larger system, where some facts lie in the heads of separate individuals while others reside in the properties of artefacts and group activities (Brown, Collins & Duguid, 1989; Sawyer & Greeno, 2009). The learning agents within the system do not primarily occupy themselves with memorising or constructing mental concepts, but are instead “getting attuned to the constraints and affordances” of the available means, methods and norms (Greeno et al, 1996, p. 17). This also implies that the possibility of separating declarative knowledge (“knowing what”) from procedural skills (“knowing how”) is impossible, and that the abstraction of practices and procedures into pure inner structures is an illusion. Hence, the transfer of knowledge from one domain into another becomes extremely difficult.

As the denomination reveals, this educational perspective mainly stems from the field of ethnography and anthropology, where researchers as Hutchins (1995) and Lave (1988) have performed comprehensive and longitudinal studies of working and learning communities. It is also influenced by activity theory and the socio-cultural school of thought, founded by Vygotsky. Important pedagogical ingredients advocated in this paradigm are meaningful and goal-oriented activities where actions could be gradually internalized, often through a step of verbalization and visualization of methods and strategies (Arievitch & Haanen, 2005). Other important aspects are the sharing of information, discussions and collaboration (Brown et al, 1989). Surprisingly, even though this pragmatic view points out the importance of appropriately designed teaching material, experimentally valid research on distributed learning in regard to such material is sparse. Schwartz and Martin (2011) have, however, shown how the constraints and affordances of mathematical manipulatives (chart pies or squares) lead to significantly different learning curves and transfer effects in fraction calculation.

Looking at all theories simultaneously, they could be regarded as separate ways of analysing different levels of cognitive aspects or properties (that is, observable individual behaviour, inner cognitive structures or contextual factors and activities at a system level). Another way of making a synthesis of the three paradigms is to consider their extension in both space and time. Then, working as a baseline and conceptual outer frame, with a long lifespan and slow changeability, is the socio-cultural context. Placed within it, we find the individuals, constantly perceiving and acting upon the surrounding world in a much quicker and responsive fashion. In the centre, inside the student’s skull, mental concepts are formed in a semi-slow pace, as the observable behaviour and explicit thoughts are evaluated and compared. And finally, serving as mediators and carriers of content between agent and context as well as between cognitive states inside the agent, are the tools, affording manipulations, visualizations or verbal explanations due to their interactive nature. It is these artefacts and their affordances that are the focus of this thesis, expected to have a measurable impact on observable behaviour and inner states. The distributed cognitive system is presented schematically in Figure 1 below:



**Figure 1:** A distributed cognitive system with mental and observable actions situated in a socio-cultural setting. The straight lines symbolise quicker processes and the serpentine slower ones.

### The role of metacognition in learning

The term *metacognition* was coined by Flavell in the 1970’s, being referred to as “*the knowledge about and regulation of one’s cognitive activities in learning processes*” (Veenman, Van Hout-Wolters & Afflerbach, 2006, p. 3). Since then, the concept has unfolded into a comprehensive research area, covering different theoretical perspectives, addressing a variety of questions and generating different kinds of studies (see Table 1 below for a taxonomy of different metacognitive components). In young children, the matter of interest generally concerns the *knowledge about one’s own and other’s mental worlds* (also called Theory-of-Mind research), and experiments often address questions about false-beliefs and the differences between mental representations and reality (Schneider, 2008). On the other hand, adults and elderly more often are studied in regard to *declarative metamemory*, where explicit beliefs about cognitive capacities as well as strategies for remembering are investigated (ibid.). Finally, *procedural metamemory*, which concerns self-regulating strategies or the capacity of evaluating one’s own acquired knowledge, often is studied with learning and developing individuals as test subjects, such as adolescents or students in various ages. Experimental designs here often contain evaluations of Feeling-of-Knowing (FOK) or Judgement-of-Learning (JOL).

**Table 1:** A categorization of metacognitive components (adapted from Schneider, 2008)

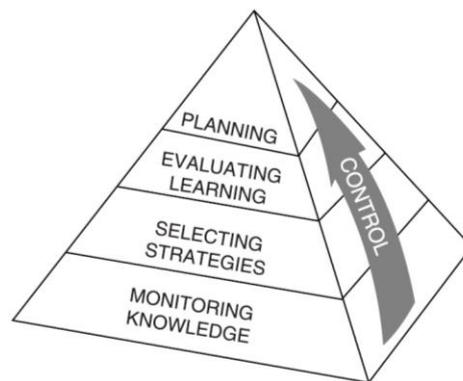
Metacognition (Knowledge of cognition)			
Knowledge about the mental world <i>“Theory-of-mind”-research</i>	Knowledge about memory		
	Declarative metamemory (variables category)	Procedural metamemory (sensitivity category)	
		Knowledge monitoring	Control and self-regulation

All theoretical fields have, in their own way, produced substantial evidence for the importance of metacognition in learning and psychological development. However, transdisciplinary research on links between the different metacognitive components - presented in table 1 - is scarce (Schneider, 2008; Veenman et al, 2006). Some of the most important findings – related to learning and school performance - are:

- Language development and early Theory-of-Mind (ToM) are linked to metamemory performance in later years (Schneider, 2008).
- Declarative metamemory as well as self-regulation and allocation of study-time improves continually during childhood and adolescence, but can be substantially different between individuals. These components can be used as trustworthy predictors of academic performance and personal development (Joseph, 2009).
- Knowledge monitoring also ameliorates during the early years, even if some studies indicate that older children do not necessarily perform better than their younger peers in FOK-tests (Schneider, 2008). The capacity of monitoring and evaluating prior knowledge is very well related to learning outcomes (Tobias & Everson, 2009).

Even though, theoretically, metacognitive components could be categorized and distinguished from one another, they also interact with and influence each other (Schneider, 2008; Veenman et al, 2006). Furthermore, since metacognitive abilities ride on and run in parallel with deeper cognitive structures that are not always explicit, they could be difficult to evaluate (ibid.). Commonly used methods within the research area of metamemory and procedural metacognition are different kinds of self-reports and self-evaluations, leading to data that is heavily reliant on the test subjects' capacity of correctly observing, judging and expressing their explicit cognitive processes and behaviour. Since the trustworthiness of such data have been proven to sometimes be low (Johansson, Hall & Sikström, 2008), methods to measure metacognition indirectly or, at least, in a less introspective manor, is of great interest (Tobias & Everson, 2009).

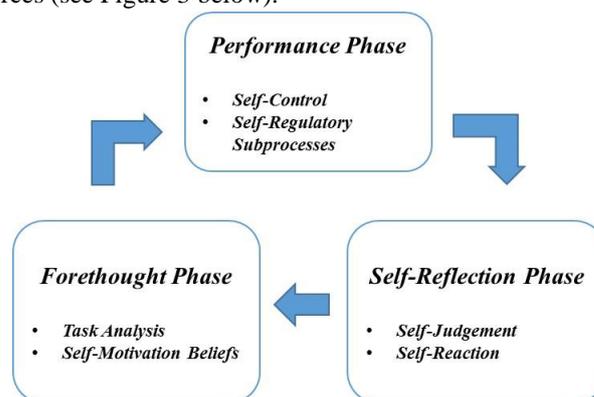
One way of studying metacognitive processes relevant to learning outcomes is to use the *knowledge monitoring assessment* framework (KMA) (Tobias & Everson, 2002; Tobias & Everson, 2009). The KMA is based on the assumption that the capacity of monitoring prior knowledge not only is essential for learning, it actually lies at the very base of the hierarchy of metacognitive processes (see Figure 2 below).



**Figure 2:** Hierarchy of metacognitive processes (Tobias & Everson, 2009).

In other words, knowing what you know and how to apply that knowledge is essential for being able to select proper strategies, to evaluate what you have learned, and to plan for further studies. Without this ability, other metacognitive processes will be less successful, and a series of studies have also shown that a correct judgement of acquired knowledge is highly related to academic performance and learning outcomes (ibid.). In contrast to many other methods of evaluating metamemory, the KMA does not rely on students reporting on how they solve problems or on what cognitive processes they use to do so. Instead, the experimental setup consists of a comparison between a feeling-of-knowing inquiry, where the test persons make judgements of whether they would be able to solve a specific problem or not, and a knowledge-related multiple-choice test. The monitoring capacity is evaluated by calculating matches and mismatches between the two tests, and the resulting coefficient has been shown to be a reliable metacognitive measure in several domains (writing, reading, math etc). This methodology has been used within the study at hand, and will be presented in detail further on.

Another, slightly more dynamic, way of addressing metacognitive aspects of learning is to regard self-regulatory processes from a socio-cognitive perspective (Zimmerman & Cleary, 2009). Since judgements, decisions and goals during different phases of the learning process depend on previous outcomes and inflict on later ones, self-regulation could be described in terms of a feedback cycle. In this model, monitoring sub-processes, norms and beliefs serve as motivational driving forces (see Figure 3 below).



**Figure 3:** self-regulatory feedback cycle with three phases (adapted from Zimmerman & Cleary, 2009).

According to this model, the *forethought phase* includes *task analysis*, where the task at hand is related to goals and strategic planning, but also *self-motivation beliefs* that are based on beliefs about self-efficacy, outcome expectations, task interest and goal orientation. All of these factors have been shown to have impact in learning outcomes, and all of them are – more or less - related to previous experiences and self-reflections.

The *performance phase* on the other hand comprises *self-controlling strategies* for reaching goals and sub-goals, such as help-seeking or time management, as well as *self-regulatory sub-processes* for successfully enhancing personal feedback. Two main sub-processes important to learning outcomes are metacognitive monitoring and record keeping, where the learning individual keeps track of his or her efforts and achievements. Several studies verify that these processes have implications for learning as well as motivation, since “tracking changes in one’s learning outcomes can produce reactive motivational effects by inspiring learners to expend greater effort” (Zimmerman & Cleary, 2009, p. 253). Notably, students tend to overestimate their self-regulating strategies and comparisons between self-reports and recorded data-logs reveal that several monitoring routines (such as revisiting figures, making notes etc.) are under-used (Winne & Jamieson-Noel, 2002; Winne & Nesbitt, 2009).

Finally, the *self-reflection phase* contains regulatory processes of *self-judgement* and *self-reaction*, both of which influence a person’s response to the previous performance. Depending on the outcome as well as the beliefs and used strategies in previous phases, a student here may react in a variety of ways. It has shown to be of great importance for successful self-evaluation not only to use controllable and explicit studying techniques, but also to apply proper goals for each step in the learning curve (Zimmerman & Cleary, 2009). Teaching strategies which have been proven to improve metacognition and thereby ought to influence the stages in this model are instructive and formative task-related feedback (Hattie & Timperley, 2007; Shute, 2007), peer learning (Schwartz, 1999), the use of teachable agents in digital learning environments (Gulz & Haake, 2015; Blair et al, 2007) and visible learning techniques with documented and displayed goals and sub-goals (Hattie et al, 2013). Several studies also indicate the need for enhancing the teachers’ metacognitive knowledge and their skills of instruction, as well as bridging the gap between the student’s knowing of their own knowledge and the teacher’s appreciation of it (Zohar, 1999). Consequently, constructing interfaces and contexts where teachers and students can meet on a metacognitive level is of great importance. As the digital tool developed in this study is meant to have an impact on self-regulation and engagement, it is designed to fit the presented model.

### *Artificial and artefactual intelligence*

So, what about the tools? To what extent could the use of sophisticated and intelligent artefacts improve metacognition and learning? During the last two decades, the professional as

well as the everyday usage of digital equipment and web 2.0 technologies have revolutionized society and the way we communicate, acquire information, manifest our knowledge and solve problems. The digitalization of learning environments has not only facilitated the distribution of material, it has also enabled distance teaching, multimodal learning applications and automatic markings of tasks and tests. In sum, computerized intelligence has a series of advantages, and it is not surprising that these benefits are both sought after and capitalized on in a variety of areas.

When it comes to education, two main types of digital and communication-related tools can be identified:

- ***Learning management systems (LMS)***, where students and teachers can receive, share and display information, create e-portfolios or participate in group activities.
- ***Instructional software***, which aim for teaching individuals or groups a specific subject or task.

The value of *LMS*, such as educational platforms or e-learning environments are of course practical, but they could also be cognitively supporting (Huffaker & Calvert, 2003; Winne et al, 2006). Following a constructivist as well as socio-cultural educational agenda, students are encouraged to not only participate in but also design and assess meaningful learning activities of their own. The use of electronic portfolios can hereby be used as a way of supporting self-regulated learning by increasing not only metacognitive awareness but also motivation and engagement (Meyer et al, 2010; Paris & Paris, 2001). However, some of the drawbacks of such systems are that they are quite time-consuming, take time to master and require extremely well-functioning technology (Trevitt, Macduff & Steed, 2014; Butler, 2006). Their impact on actual measurable learning outcomes amongst unengaged students is also a matter of discussion (Abrami et al, 2013). Still, without applying a totally constructivist approach to learning, nothing prevents *LMS*’s to support more instructional and regulated teaching methods. These systems could, for instance, improve learning by facilitating formative feedback, individualizing instructions and tasks or visualizing learning processes in relation to goals and agendas. The collection of logs and activities from such systems equally makes it possible to analyse and evaluate the “real” use of well-known strategies, leading to better knowledge about learning styles, socio-cultural or neuro-cognitive conditions and their impact on self-regulation (Winne & Jamieson-Noel, 2002).

One additional important factor when sharing information between individuals in a heterogenous crowd (in this case teachers, students, parents, school principals etc.) is correct transformation and presentation of data for each group of users. As researchers studying interactive and cooperative software often point out, information is created in a *context*, for a specific purpose in mind (Dourish et al, 1993, Ackerman et al, 2013). Sharing it outside its original environment, making it receivable for a larger audience, is often technically simple, effective and time saving. The issue for the “outside” receivers

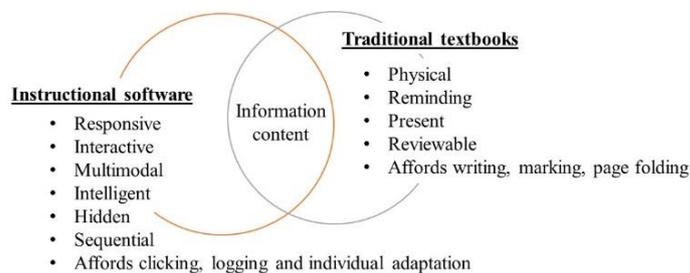
is here to be able to correctly *interpret* the meaning and purpose of the information (ibid.). This aspect, and the need for adaptation and reconstruction of data within a digital network with many users, is frequently studied and discussed in professional environments with a high amount of shared information – especially in hospitals or other medical settings (ibid.). Similar research within LMS-systems are hard to find. It is however very likely that the specific informational content shared between educational staff (concerning goals or learning progress, for instance) is not always suitable for communication between parents and teachers or teachers and students, and so forth. In other words, visualizing knowledge is not a one-way street, and research on appropriate interfaces in digital learning environments is lacking.

The other important group of learning technologies consists of *instructional software*, i.e. digital applications in various shapes designed for (with or without the help of assisting teachers) transferring knowledge in a specific domain (math, writing, reading, science etc.). Even if these applications today are hard to categorize due to their heterogenous and multifunctional nature, they normally contain one or several of the following educational features: *Drill- and practice*, *Tutorial*, *Simulation*, *Instructional game* and *Problem solving* (Doering & Veletsianos, 2009, Sjödén, 2015). While the first two of these are compatible with more traditional learning theories within the behaviouristic tradition, the following three could be applied within any of the theoretical frameworks described in earlier passages. Central advantages of instructional software are the possibility of individualization, multimodal interaction and immediate feedback. By constructing branches with different content or various degrees of difficulty, students with different learning styles and different knowledge levels can work within the same subject according to their personal preferences and preconditions. Tasks and topics can be learned through exploration and by following game-like rules, and virtual worlds can be created and peopled with avatars and agents, suitable for role-play or instruction.

The possibilities with instructional software appear to be endless, but the supply of applications that keep their promise and really leads to deep learning or metacognitive awareness and motivation are few, at least in relation to the abundance of apps and games available on the market (Sjödén, 2015; Gulz & Haake, 2015). Some frequently debated issues concerning the pedagogical value in such products is the risk for tempting the user to apply a trial-and-error-like behaviour, not paying attention to feedback and not relating mistakes or successful outcomes to specific knowledge or informational content. Some of these matters can be addressed and controlled for by careful design and testing (the way children use instructional software could be quite different from the way it originally was intended) and others can be monitored or avoided by involvement from teachers or peers (Doering & Veletsianos, 2009; Sjödén, 2015; Lindström et al, 2011).

One rarely discussed matter regarding instructional software is, however, the actual affordances and properties the digital media possess, and how those deviate from the af-

fordances of more traditional and paper-based material. Remembering the distributed cognitive system presented in Figure 1, and the fact that learning takes place in a context where the student gets attuned to the constraints and affordances of the material at hand, these differences ought to be of interest. A set of examples of such discrepancies is therefore presented in Figure 4 below:



**Figure 4:** A comparison between traditional textbooks and instructional software in terms of affordances and physical properties.

One major difference between a book and a digital application is their extension in space and the way they are physically manipulated and perceived. A book, no matter how static and unintelligent, does not disappear by a click. It can lie in front of you as a reminder, of what you have read, ought to read, have learned or will learn. By the size or shape of the book you can estimate its extent of information, you can revisit earlier passages and remind you of earlier performed tasks. Objects equally permit manipulation, and physical artefacts gain intelligence by their physical entities, by being movable, twistable, stackable or groupable, thereby facilitating searching and problem solving (de Léon, 2003). Studying human cognition in a work-related or educational context, it also becomes clear that our limited memory capacity makes us reliant on external memories and placeholders (Dix, Ramduny-Ellis & Wilkinson, 2001; Kirsh, 2004; Wilson & Clark, 2009). Short notes on a piece of paper, an unfolded paper file or a well-placed item may inform us of where we last were, what we were doing or what we are about to do. Since metacognition rides on cognitive processes (as we have stated before), the need for digital functionality to support these matters is, of course, of great interest.

Another aspect of artefacts is their capacity for creating a sense of ownership. Many school books are personal and owned by the students, with names or perhaps personal tags written on the front cover. This marking might be seen as a sign of possession (it is only me that has the authority to write and read in this book), but could equally be interpreted as a sign of *psychological ownership* of the content (in this book lie my achievements and my knowledge, my physical proof of my days at school). Psychological ownership towards one's work is often discussed as a major factor of success within organisations and companies (Avey et al, 2009; Pierce, Kostova & Dirks, 2003). Without assigning to much invaluable properties to old-fashioned learning material, these aspects ought to be addressed when instructions and tasks become digitalized.

Surprisingly, in contrast to the LMS described in the passages above, instructional software is rarely evaluated in regard to its way of visualizing learning progress or achieved knowledge as an instrument for engagement and metacognitive awareness. The importance of logging accurate data is often emphasized, as this is valuable for teachers as a tool for assessment (Gulz & Haake, 2015; Sjöden, 2015), but such data is not always accessible - or often extremely quantitatively and statistically summarized - for the students themselves.

In games on the other hand, award-like features like badges or stars, as well as the access to richer worlds or more advanced levels, often are used as proofs of achievement. These normally serve as motivating tokens, but they are rarely related to the specific content of the acquired knowledge or skills. In the figures below, two examples of visual cues for interpreting learning outcomes in two different types of software are presented.

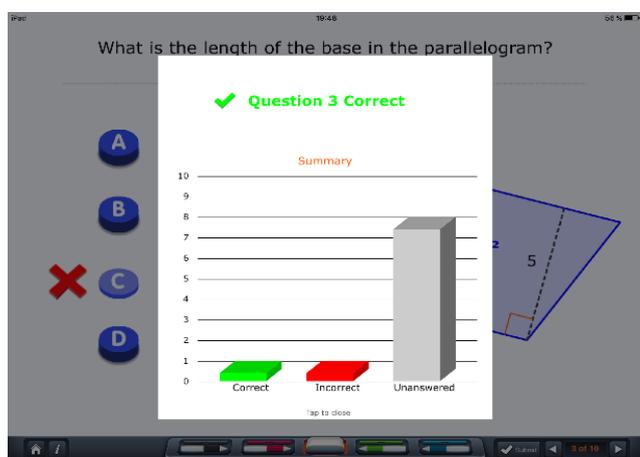


Figure 5: A geometry application, showing the quantity of correct, incorrect and still unanswered tasks (Area of Figures, 2014).

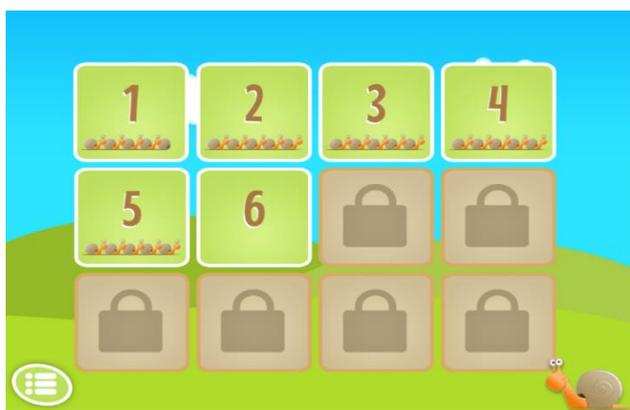


Figure 6: A spelling game with snails and levels as rewards and proof of achievement (Happi Stavar, 2014)

### Guardians of History

Learning management systems or instructional software are not only useful as tools for instruction or exploration, they can also be utilised as powerful research instruments (Gulz & Haake, 2015). The Educational Technology Group (ETG) at the universities of Lund and Linköping in Sweden have developed and used learning applications as a means for inquiry and research during the last decades, leading to important findings

regarding children's learning processes and their actual use of digital software in realistic class-room experiments (Axelsson, Andersson & Gulz, 2016; Gulz & Haake, 2015; Lindström et al, 2011).

One of these applications is *Guardians of History (GoH)*, a software which aims to teach history of science and source criticism to students in 4<sup>th</sup> to 6<sup>th</sup> grade using explorative and tutorial content mixed with drill-and-practice tasks in a game-like fashion. The students work individually and are given missions in which they are told to perform time-travels to different historical persons and events, to explore these environments and search for information. The software makes use of a Teachable Agent (TA), why the student after his or her time-travel returns to the original narrative setting – a castle with elves with different characters and roles - and teach one of these elves historical facts (for a detailed description of the TA-paradigm, see for instance Blair et al, 2007). The design of such a teaching task varies, from concept maps to multiple-choice tests and narratives with missing words. After the elf has been taught it writes an essay on the topic, and depending on the outcome of this task the mission is approved or rejected. A successful mission leads to new missions and travels, while a failed one means that the elf must be taught all over again. The historical destinations are saved and can be revisited an infinite number of times, also after the mission has succeeded. Figure 7 below presents the structure of the application with its paths and conditions.

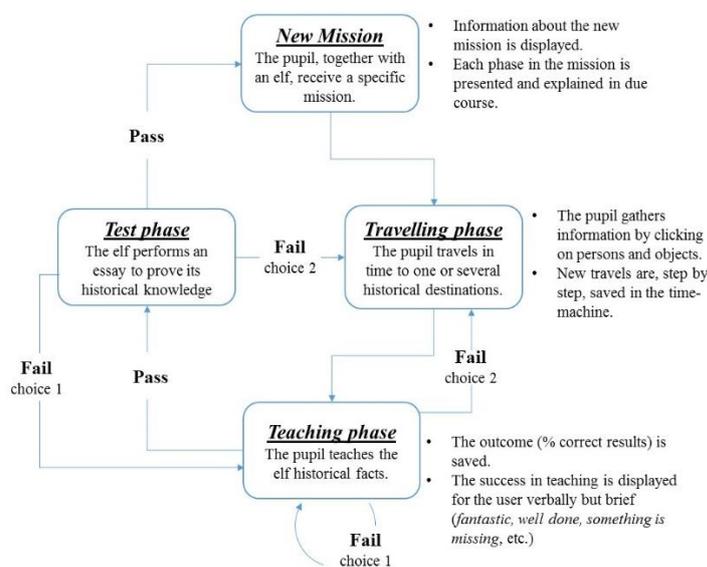


Figure 7: The conceptual structure of Guardians of History

GoH can be set up with a stack of different missions and tasks. As used in the present study, the introductory assignment concerns source criticism, where the student mainly performs tasks with instructions from an elf-archivist in the castle archive (see Figure 8 below). Other historical topics are the industrialism, the plague in London during the 17<sup>th</sup> century, and scientists like Galilei, Newton and Emilie de Chatelet.



**Figure 8:** A view of the cellar archive in the castle, with the elf “Nauvo” as a pedagogical agent.

As Figure 7 shows, GoH is mainly linear and sequential, and when the player begins he or she neither knows how many missions the application contains nor their corresponding travel destinations. Each new mission is presented when the former one is terminated, and the only visualized history about the student’s accomplishments is in terms of old time-travels saved in the time-machine. Data-logs are of course saved for professional analysis and evaluations, but these are neither suitable nor accessible for the user during play.

In sum, GoH represents a common type of learning application where the student strives for progress by exploring and gathering information and performing knowledge-related tasks. It affords mainly clicking, hovering and reading, even if some short instructions are available on video. During play, the user can make notes - on paper or within a function inside the software.

### *Hypotheses*

Based on the theoretical and practical facts presented in the passages above, there is substantial evidence that tools and artefacts can implicate learning processes and metacognition in a variety of ways, not only depending on their content, but also due to their specific affordances. By visualizing learning progress and presenting scaffolding feedback on personal results, there is a good chance that self-regulating strategies as well as knowledge monitoring could be facilitated. At the same time, the digitalisation of teaching aids and educational tools steadily increases, and some schools also actively try to avoid traditional paper based material. Still, studies on instructional software and their possible need for metacognitive elements are rare. Hence, the aim of this thesis is to fill in this research gap by comparing two versions of an instructional software, *Guardians of History*, where one of them has no visual real-time presentation of performance or acquired knowledge, while the other one does. (For displaying this information, an interactive *visualization tool* has been designed, see passages below for further description). The following hypotheses have

been formulated to address the issues of metacognitive impact of such knowledge-related visualizations:

H1: A digital function, displaying visual information about the user’s performance and knowledge, will be interacted with, indicating that it is attended to and therefore cognitively processed by the users.

H2: The interaction with such a function - and its content - will improve metacognitive processes, i.e. knowledge monitoring.

H3: The function itself and the possibility of gathering tokens will have a positive impact on self-regulation, by supporting endurance and engagement.

The upgraded *Guardians of History* could here be regarded as a mix of LMS and instructional software, exploring issues relevant to both groups of tools. In addition to the specific hypotheses, the study aims to explore and discuss user patterns and user preferences that could be related to the stated research questions. The experimental setup for this research agenda rests on a multiple-method design, which is presented in further detail in the passages below together with a more detailed description of the used stimuli.

## **2 Method**

The hypotheses were addressed by gathering quantitative as well as qualitative data in an experimental setting with high ecological validity. Furthermore, as the passages below will reveal, the entire field-work was conducted in cooperation with another ETG-study on learning and motivation<sup>1</sup>. Consequently, all participants were subjects of two inquiries at the same time, sharing varieties of two stimuli categories (one for each project) and performing activities and tests relevant for both studies.

### *Participants*

Five classes (three in the 5<sup>th</sup> grade and two in the 6<sup>th</sup> grade) were recruited from two Swedish primary schools. The schools and classes were selected by convenience sampling through the help of university staff and their contacts. In total, 117 students between 11 and 12 years old (61 girls and 56 boys) participated in the study. The results from five participants were excluded from the dataset, either due to insufficient gameplay (at least one mission had to be accomplished) or due to absence during final tests. Although the students’ attendance in this case fell under the teachers’ consent, the students as well as their parents were given thorough information about the purpose and content of the study. They were also given an opportunity to decline any use of their data, a choice no one actually made. After collection and digitalisation, all data were anonymised. The teachers did not get access to any information about the students’ individual achievements or results.

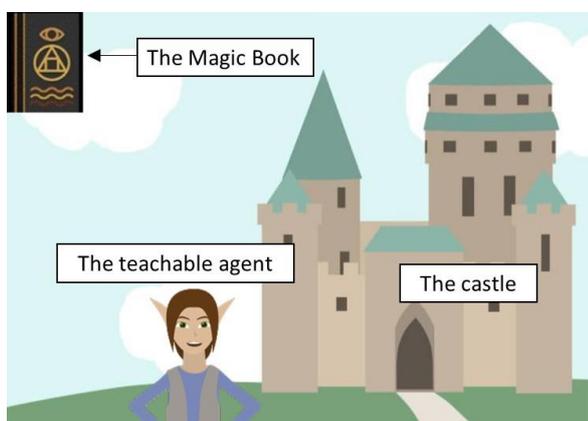
All children received a personal diploma as a gesture of gratitude for participating. Four of the classes also accepted an

<sup>1</sup> Designed and performed by Agata Janiszewska

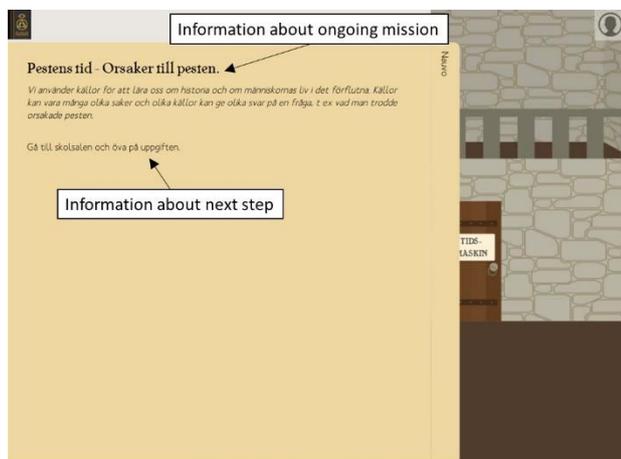
offer to visit the university and two of its laboratories (the motion capture lab and robotics lab).

### Materials

The stimuli for the study at hand consisted of the instructional software Guardians of History (described briefly in the passages above) containing 8 solvable time missions (Source-criticism, Industrialism 1, Industrialism 2, The Plague in London, Sophie Brahe, Galileo Galilei, Isaac Newton and Emelie de Chatelet). The game also contained two versions of a clickable feature, called “the Magic Book” (see Figure 9 below), with different content and functionality. In the baseline condition, the content of the book was limited to descriptions of ongoing missions and time-travels, and the student could click on it in case of uncertainty about what to do or where to go next (see Figure 10).



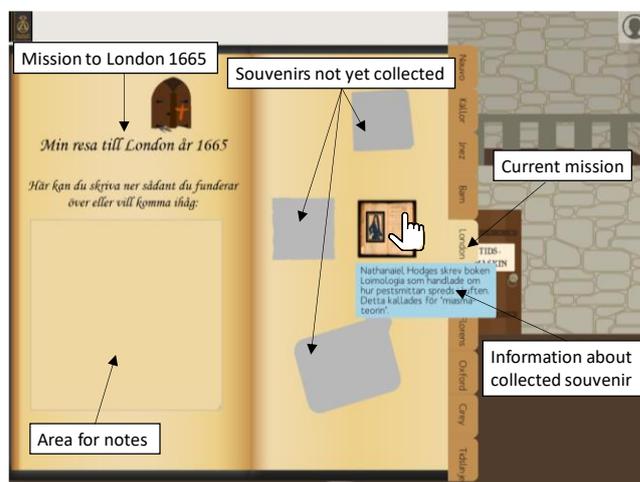
**Figure 9:** A view of the castle, with the teachable agent (one of the elves) and the clickable magic book.



**Figure 10:** Baseline condition (type A) of the Magic Book with instructions about actual mission (the plague in London), and the next step in the game (in this case, to go to the classroom and teach the elf).

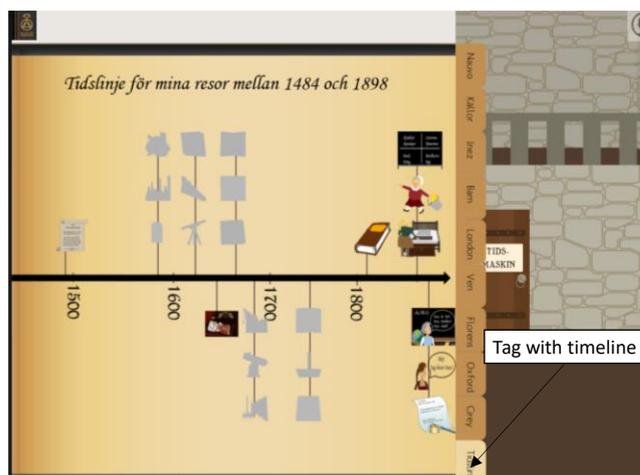
In the extended version of the book, the user also received time-travel related souvenirs for solved tasks. Each mission was here assigned to a specific tag in the book, giving the player an opportunity to reflect upon acquired knowledge in relation to the actual mission, but also to study how many missions that were up ahead and what to expect next. The souvenirs could be hovered, (by holding the fingertip or the cursor

on them), then showing compressed information from the related tasks (see Figure 11). This function added a sort of step-wise feedback to the game, making it possible for the player not only to attend to visual proof for partly solved missions but also to repeat learnt topics without travelling back in time all over again. The area to the left of the souvenirs was dedicated to notes which could be written at any time before, after or during time-travelling, contributing to a notion of travel diary to the user.



**Figure 11:** Experimental condition (type B) of the Magic Book with one gathered souvenir for one partly solved mission (about the miasma-theory during the plague in London 1665).

After gathering all souvenirs from one mission, they could be transferred to a timeline (see figure 12 below). To catch the user’s attention, the book was introduced (although rather briefly) in the instructing dialogue within the game, and the book blinked anytime a new souvenir was added.



**Figure 12:** Type B of the Magic Book with gathered souvenirs put on a prepared time line.

The stimuli for the concurrent ETG-project consisted of different versions of one of the characters in the game and did not affect the material presented above. That study also made use of a kind of collectables, but these could not be mistaken for the tokens relevant to the study at hand.

Note that the Magic Book - as well as some features for the concurrent study and certain aspects of the application in general - was specifically implemented for the present study.

Consequently, to verify that the software was usable and free of bugs, it was tested on a separate group of 5<sup>th</sup> graders approximately one month before the regular experiments. This trial resulted in minor corrections in code but did not lead to any alterations of the design and content of the stimulus. The GoH software was put on an external server (Heroku.com), accessible through the internet by approved logins. All player activity (clicks on clickable objects etc.) was registered on this server with timestamps (UNIX time), identification codes and user configurations. This data was used to identify, explore and evaluate dependant variables for the first and third hypotheses stated above.

For measuring knowledge monitoring (the second hypotheses), a two-part KMA-test was constructed (Tobias & Everson, 2002; Tobias & Everson, 2009). The first part consisted of an FOK-test with specific knowledge statements with yes- and no-answers (such as “*I know where the plague spread in 1665.*”). The second part consisted of a knowledge test with both open and multiple-choice questions (such as “*Where did the plague spread in 1665?*”). The two tests were related so that the content overlapped and could be compared. Since the second test also was used in the study on learning and motivation, not all questions were relevant to the study at hand. The two tests are presented in full in Appendix A and B where the relation between them also is described. Note that some of the questions were not included in the final calculations. The reasons for this is further explained in the result chapter below.

Measuring knowledge monitoring by this kind of related tests has been shown to produce indexes that correlates well to performance and metacognitive capability (ibid.). Still, research also show that the results from KMA-tests are more correlated with these capacities when evaluating general prior knowledge than domain-specific know-how and expertise (ibid.). This methodology has, to the authors knowledge, not been used in any study evaluating the effects of instructional software. Neither has it been used for evaluating short-term effects on different learning strategies within a specific domain. The advantages of the method are mainly its quantitative nature together with the minimal element of self-reporting and explicit self-evaluation. Still, the use of KMA within this study should be considered as both exploratory and unpredictable.

Finally, a short qualitative questionnaire with open ended questions regarding the users’ experience and general game design was constructed (see Appendix C). These questions were mainly created to obtain the participants’ subjective opinions and to receive ideas on possible game development. However, they have also been utilized – together with qualitative findings from group discussions in class - for shedding light on user preferences regarding metacognition and digital learning.

### Experimental design

The experiment for the present study was based on a between-subjects design with two varieties of the magic book as independent variables - Type A, with no souvenirs and only the help function, and Type B, with both the souvenirs and the help. Since the simultaneous study for its part used three independent conditions, each class was divided into six groups containing all combinations of the different stimuli configurations (two times three conditions).

To ensure that students with different performance levels were evenly distributed throughout the groups, the teachers were asked to rate the children’s reading performance on a three-dimensional scale (high, medium and low)<sup>2</sup>. The table below shows the conceptual construction of the six groups in a hypothetical class with 24 students (6 low performers, 6 high performers and 12 medium performers). To facilitate analysis and data tracking, the students were given login and passwords that corresponded to their designated stimuli configuration.

**Table 2:** Experimental design for one hypothetical class with 24 students.

	ETG-study 2, condition 1	ETG-study 2, condition 2	ETG-study 2, condition 3
Magic Book, Type A	<u>Group 1A</u> 1 L perf. 2 M perf. 1 H perf.	<u>Group 2A</u> 1 L perf. 2 M perf. 1 H perf.	<u>Group 3A</u> 1 L perf. 2 M perf. 1 H perf.
Magic Book, Type B	<u>Group 1B</u> 1 L perf. 2 M perf. 1 H perf.	<u>Group 2B</u> 1 L perf. 2 M perf. 1 H perf.	<u>Group 3B</u> 1 L perf. 2 M perf. 1 H perf.

Each one of the classes committed three full lessons, between 45 minutes and one hour long, to the study. Due to breaks, holidays and other school activities, the time between the three experimental sessions (one for each lesson) varied between the classes (see figure 13 below). While the first two sessions were devoted to game play, the third was dedicated to tests and discussions. Between the second and third session, half an hour was designated to play-time at home.

<sup>2</sup> Variables originally designed for the concurrent ETG-study by Agata Janiszewska. Used with permission in the study at hand.

	Class 5.1	Class 5.2	Class 6.1	Class 6.2	Class 5.3
Day 1	Session 1	Session 1			
Day 2	Session 2	Session 2	Session 1	Session 1	
Day 3					
Day 4	Home play	Home play			
Day 5					Session 1
Day 6	<i>Spring break</i>				
Day 7					
Day 8					
Day 9					
Day 10					
Day 11					
Day 12					
Day 13					
Day 14					
Day 15					
Day 16	Session 3	Session 3	Session 2	Session 2	
Day 17			Home play	Home play	
Day 18			Session 3	Session 3	
Day 19					
Day 20					
Day 21					
Day 22					
Day 23					Session 2
Day 24					Home play
Day 25					
Day 26					Session 3

**Figure 13:** Experimental sessions for each one of the five classes

The three main dependant variables chosen for testing the hypotheses stated above were the time spent in the Magic Book (H1), the KMA-index from the FOK-questionnaire and knowledge test (H2), and the time played at home (H3). These variables interrelate to the distributed cognitive system and its various levels and processes in the following way:

**Table 3:** The measured variables and their place in the cognitive system

Variable	Data	Level of analysis	Cognitive processes
<i>Time spent in Magic Book</i>	Data Logs	Behavioural	Attention
<i>KMA-index</i>	Hamann coefficient	Mental	Knowing of knowing
<i>Time played at home</i>	Data Logs	Behavioural	Executive functions & Self-control

In addition to these measurements, any qualitative data obtained from the students' own verbal reports and expressions could be regarded as both mental and sociocultural constructs. A more detailed description of the experimental procedure and the variable calculations is presented further down in the text at hand.

### Procedure

Before starting the first session, the students were introduced to the research area and the instructional software through a 30-minutes presentation (in class or at the university). Each of the two first lessons then started with a brief résumé of what

to do, what to expect and what to bear in mind while playing. The students were encouraged to ask for help if needed and to take notes on the side during the session. They were also empowered to read all instructions in the game (mainly presented through dialogues) thoroughly and to prepare themselves for a considerable amount of text with historical content. To avoid associations with traditional computer games the instructional software was mainly referred to as "the learning material", and the term "game" was deliberately avoided. The participants were not informed that they were submitted to six different variants of the game, but they were not prohibited to talk to each other and to discuss their personal experiences.

After the second session, the students were given a choice: to continue to play at home for at least half an hour or to do an alternative homework (picked out by the teacher). During the third and last session, they filled in the questionnaires (FOK-test, knowledge test, report on user experience and an additional inquiry on motivation for the concurrent study). All tests were distributed on paper. Subsequently, the FOK-test was designed with one question per sheet, and the students were instructed to turn the page on an oral command, giving them approximately 10-12 seconds to reflect on each statement. The time for the knowledge test was limited to 25 minutes, and the user experiences reported after this was finished in case of spare time.

Before taking leave of the children, the true nature of the experimental setup was revealed. In three of the five classes, there were also time for further questions, feed-back and group discussions. All sessions were conducted in ordinary class rooms under the lead of the authors of this and the concurrent study. All participants, independent of assigned condition, received the same instructions and support.

## 3 Results

In total, datasets from 112 students - 61 within condition Type A and 51 from condition Type B - were used for analysis and evaluation. The amount of data in each group is hereby considered to be statistically satisfactory (see for instance Van Voorhis & Morgan, 2007, for rules of thumb regarding appropriate sample sizes). However, due to ecological nature of the experiment in hand, the actual conditions and practicalities were not entirely equal between experimental sessions. Factors that couldn't be fully controlled for and therefore differed from class to class were the capacity of the wireless network, used hardware (tablets and computers) and software (operating systems and web-browsers). Some students also suffered from occasional software errors, and the need for assistance and instruction varied greatly between classes, sessions and individuals. The possibility of exceeding the time in class - in case of technical problems - also varied from session to session, leading to reduced playing time for some students. Subsequently, to ensure that the two groups are fully comparable, they have been reviewed before statistically testing the stated hypotheses and evaluating additional qualitative findings.

All statistical analysis has been performed with the statistical software R version 3.2.4 (R Core Team, 2016) at an alpha level of 0.05 if not otherwise stated.

### Group control

The two conditions in GoH are not supposed to lead to different success-rates or different numbers of solved tasks and missions. And, since the progression in the game has an impact on the use of the Magic Book as well as on metacognitive measures, the proportion of students solving different numbers of missions is compared between groups (see Figure 14 below). The distributions for the two conditions are not normally distributed, as per Shapiro-Wilks test ( $W = 0.81, p < 0.001$  for Type A and  $W = 0.91, p = .001$  for Type B). Consequently, any differences have been evaluated through non-parametric methods.

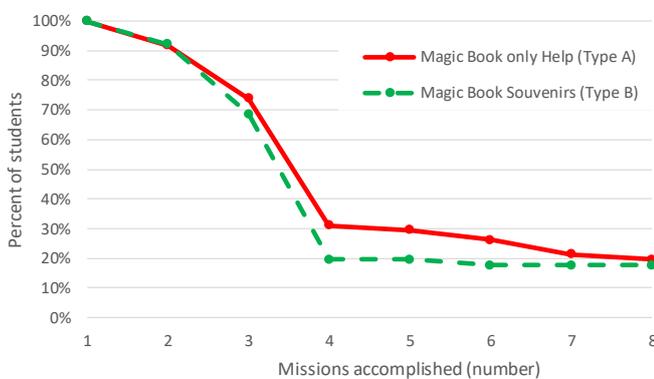


Figure 14: Distributions for solved missions.

Although it might appear as the participants in the Type B condition have solved slightly fewer missions, a Mann-Whitney's U-test shows that the difference between groups is not significant ( $W = 1711, p = .34$ ). As the diagram reveals, most of the students succeeded in solving the first to third mission, while the fourth was considerably harder, separating unmotivated or less capable students from the more ambitious and high performing ones (note that the players solved missions sequentially, see Figure 7).

Another factor, influencing progression as well as other variables, is the actual playtime in school for the students within each group. Since some classes had more disturbances than others, this aspect has also been evaluated (see Figure 15):

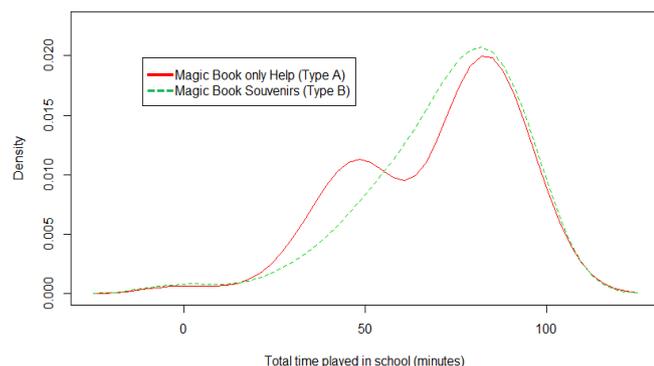


Figure 15: Distributions for playtime in school.

The distributions are not normally distributed, as per Shapiro- Wilk test ( $W = 0.91, p < .001$  for Type A and  $W = 0.90, p < .001$  for Type B). A Mann-Whitney's U-test reveals no significant differences between conditions ( $W = 1454, p = .56$ ).

Finally, to assure that the performance levels of the students does not differ between groups, the distributions of low, medium and high performers (regarding reading capability) also have been compared (see Table 4 below). No significant differences can be found, as per Mann-Whitney's U-test ( $W = 1735, p = .17$ ).

**Table 4:** The number of students with different reading capabilities within each group.

	Performance level reading		
	Low	Medium	High
Magic Book only Help (Type A)	19	17	25
Magic Book Souvenirs (Type B)	16	13	22

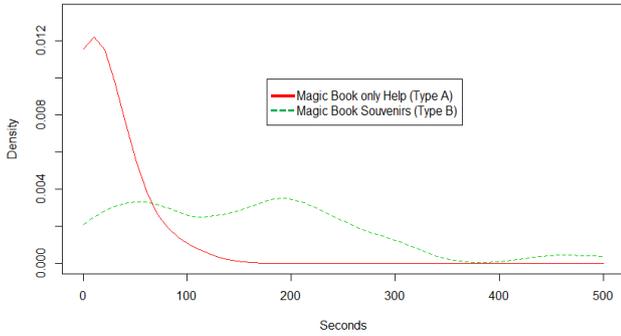
The lack of significant differences in the comparisons above implicates that the two groups contain students with equal strengths and capabilities and that they have been exposed to the material under equivalent time.

*Hypothesis 1: A digital function, displaying visual information about the user's performance and knowledge, will be interacted with and attended to.*

As dependant variable for testing the hypotheses, the time spent (in seconds) on clicking and hovering on artefacts in the magic book has been used. Consequently, this time factor is assumed to be an appropriate measure for attention and perceptual processing of the graphical and verbal content. Since both groups had access to the clickable book, the main difference between the two conditions was the content of the actual stimuli. A minor difference was also that the book in the Type B condition (with souvenirs) blinked as soon as the content was updated.

To reassure that the logged timestamps reflects proper interaction, the attention towards a clicked and visualized object is supposed to never last longer than three minutes (180 seconds). Longer sequences between clicks are assumed to be due to the user ignoring the game or perhaps even leaving the computer. Note that a substantial amount of play-time was effectuated at home, without any observation of the users' behaviour.

The distributions for the time spent in the Magic Book in the two conditions are presented in Figure 16. Neither of these populations are normally distributed, as per Shapiro-Wilks test ( $W = 0.73, p < 0.001$  for Type A and  $W = 0.89, p < 0.001$  for Type B). A Mann-Whitney's U-test reveals a significant difference in spent seconds between Type A ( $Mdn = 10, Range = 0-105$ ) and Type B ( $Mdn = 158, Range = 0-607$ );  $W = 286, p < 0.001$ .



**Figure 16:** Distributions for spent seconds in the Magic Book.

*Hypothesis 2: The interaction with the digital function - and its content - will have a positive impact on metacognitive processes, i.e. improve knowledge monitoring.*

The dependant variable for addressing this issue is the *Hamann coefficient*, (*HC*), calculated between the two parts of the KMA-test (the meta-test, attached in Appendix A and the knowledge test, Attached in appendix B). *HC* is a well-known accuracy index that has been used in multiple studies on the capacity of monitoring prior knowledge (Tobias & Everson, 2002; Tobias & Everson, 2009; Wright, 1996). It is calculated by comparing the amount of mismatches and matches between tests according to the following procedure: First, the FOK-answer in the meta-test is compared with the answer in the knowledge test. This is done question per question for each participant (see Table 5).

**Table 5:** A 2 x 2 contingency table showing recognition performance and feeling of knowing judgements. The letters, a through d, refer to the cell frequencies (adapted from Wright, 1996).

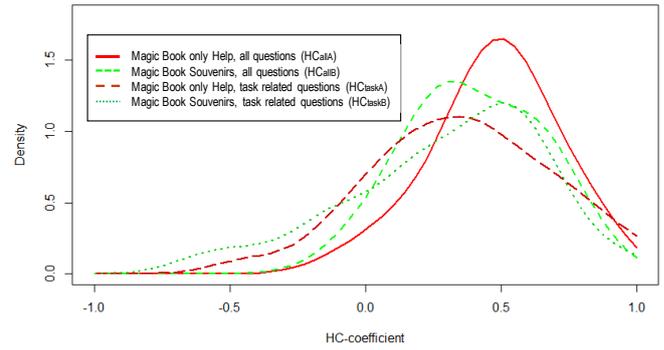
		Recognizing (according to the knowledge test)	
		Yes (correct)	No (incorrect)
Feeling of knowing	Yes	a	b
	No	c	d
Matches:		a+d	
Mismatches:		b+c	
n:		a+b+c+d	

The Hamann coefficient (*HC*) is then calculated through the equation below, leading to a number between -1 and 1 for each participant:

$$HC = \frac{(a + d) - (b + c)}{(a + b + c + d)}$$

Since the *HC*-coefficients are affected by the student's progression (solving more tasks in the game ought to influence response-rates and thereby the proportion of correct answers), the values have been calculated in two different ways to for each of the two conditions: For all questions in the test, regardless of the student's own progression and the number of solved missions (*HC<sub>all</sub>*), and for those questions representing the student's actual solved missions and tasks (*HC<sub>task</sub>*). The distributions for both *HC<sub>all</sub>* and *HC<sub>task</sub>* are presented in Figure 17. These could not be dismissed as not normally distributed, as per Shapiro Wilks test ( $W = 0.97, p = .10$  for *HC<sub>allA</sub>* and  $W = 0.97, p = .23$  for *HC<sub>allB</sub>*,  $W = 0.98, p = .53$  for *HC<sub>taskA</sub>* and  $W = 0.96, p = .055$  for *HC<sub>taskB</sub>*). Consequently, parametric statistics can be used for analysing *HC*-coefficients in relation to explanatory variables. Means and standard deviations for the measures are presented in Table 6.

$W = 0.96, p = .055$  for *HC<sub>taskB</sub>*). Consequently, parametric statistics can be used for analysing *HC*-coefficients in relation to explanatory variables. Means and standard deviations for the measures are presented in Table 6.

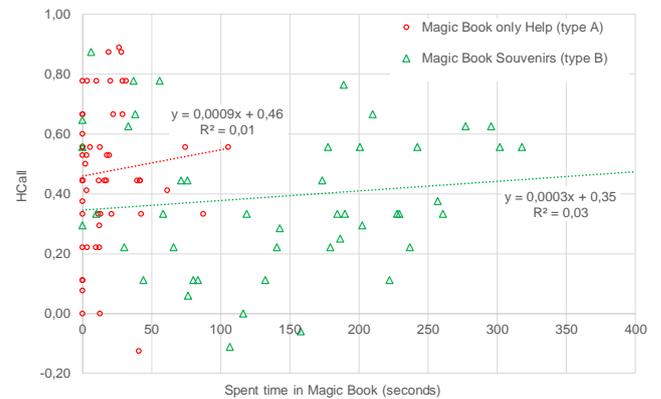


**Figure 17:** Distributions for *HC*-coefficients

**Table 6:** Results of the Hamann-coefficient calculations

	<i>HC<sub>task</sub></i> (M, SD)	<i>HC<sub>all</sub></i> (M, SD)
Magic Book only Help (Type A)	0.36, 0.33	0.47, 0.22
Magic Book Souvenirs (Type B)	0.31, 0.35	0.40, 0.24

Since the performance and progression has been evaluated as equivalent in both groups, and a significant difference in use of the Magic Bok has been found, the *HC*-coefficients ought to be possible to compare through traditional independent T-test-calculations. However, a scatterplot reveals a possible correlation between the *HC*-value and the amount of individually spent time in the Magic Book, indicating a need for a more complex analysis:



**Figure 18:** Spent time in Magic Book (for single students) plotted against *HC<sub>all</sub>*.

In addition, general performance-levels are known to influence KMA-indexes and therefore risk masking a small but significant Magic Book-impact. Consequently, the following three explanatory variables have been selected for fitting *HC*-coefficients into a linear model:

- 1) Reading performance (*perf*)
- 2) Magic Book Type (*mbType*)
- 3) Spent time in the Magic book (*mbTime*)

A multiple regression analysis with control for interaction effects has then been performed according to the following formula:

$$HC_{(task, all)} = A + \beta_1*perf + \beta_2*mbType + \beta_3*mbTime + \beta_4*mbType*mbTime + \beta_5*perf*mbType$$

The analysis results in a significant regression equation for both  $HC_{all}$  ( $F(5,106) = 5.9, p < .001$ ) and  $HC_{task}$  ( $F(5,106) = 3.4, p = .01$ ), where  $perf$  and  $mbType$  are the only variables with significant effect on the calculated HC-values. As expected,  $perf$  has been found to be a significant predictor for both  $HC_{all}$  ( $\beta = 0.44, p < .001$ ) and  $HC_{task}$  ( $\beta = 0.30, p = .01$ ). A more unexpected effect was that  $mbType$  demonstrated a small but significant **negative** effect on  $HC_{all}$  ( $\beta = -0.28, p = .036$ ). This implies that the Type B version of the book could lead to diminished metacognitive monitoring – quite opposite to the stated hypotheses. Table 7 summarizes the results for all variables:

**Table 7:** Multiple regression analysis for  $HC_{all}$  and  $HC_{task}$

	$HC_{task}$			$HC_{all}$		
	Estimate	Std. Error	$p$	Estimate	Std. Error	$p$
Intercept	0.07	0.11	.55	0.22	0.07	.003*
$perf$	$\beta = 0.30$	<b>0.05</b>	<b>.01*</b>	$\beta = 0.44$	<b>0.03</b>	<b>&lt; .001*</b>
$mbType$	$\beta = -0.11$	0.10	.44	$\beta = -0.28$	<b>0.06</b>	<b>.036*</b>
$mbTime$	$\beta = 0.53$	0.002	.42	$\beta = 0.31$	0.001	.62
$mbType*mbTime$	$\beta = -0.004$	0.002	.39	$\beta = -0.0004$	0.001	.92
$perf*mbType$	$\beta = 0.0002$	0.0004	.71	$\beta = -0.0001$	0.0002	.77

$$F(5, 106) = 3.4, p = 0.01 \quad \left| \quad F(5,106) = 5.9, p < .001 \right.$$

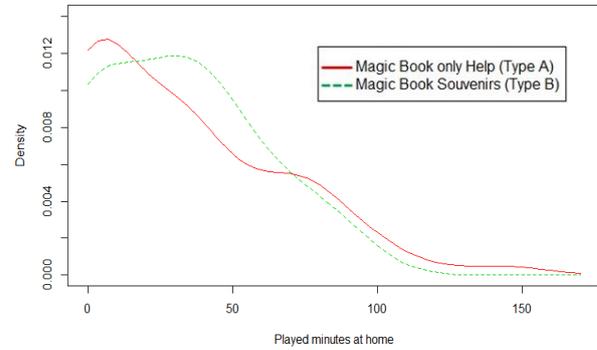
$$R^2 (adj) = .10 \quad \left| \quad R^2 (adj) = .18 \right.$$

*Hypothesis 3: The function itself and the possibility of gathering tokens will have a positive impact on self-regulation, by supporting endurance and engagement*

A factor describing self-control and metacognitive monitoring is the allocation of study time. Winne and Nesbitt (2006) here point out that this aspect of learning differs between low- and high performing students, and that many students are in need for tools and support in this area. One might also claim that the longevity of voluntary gameplay using a digital learning material could reflect a kind of engagement and endurance, especially if the student progresses slowly and he or she finds the material challenging or difficult.

As dependant variable for testing the hypothesis, the student's amount of playtime at home has been selected. Note that the task of playing at home was possible to decline in favor for another kind of homework selected by the teacher<sup>3</sup>. According to the teachers, none of the students actually chose the other homework, even though the logs reveal that several students didn't play at home during sufficient time (34 of the students only played for a minute or so at home, while 20 of them restricted their digital homework to last 2 to 29 minutes). Naturally, any student reaching the end of the game within the

dedicated 30 minutes could impossibly continue playing, why the data for such users has been eliminated from the dataset. In sum, data from 57 players in the Type A condition and 49 in the Type B condition has been used. The distributions for the playtime at home for the two groups are presented in Figure 19 below. Neither of these populations are normally distributed, as per Shapiro-Wilks test ( $W = 0.86, p < .001$  for Type A and  $W = 0.91, p = .001$  for Type B). A Mann-Whitney's U-test ( $W = 1397, p = 1$ ) reveals no significant differences between conditions.



**Figure 19:** Distributions of played time at home for the two conditions.

*Additional findings 1: Specific use of the visualization tool*

Although the first and second hypotheses have been statistically evaluated in the passages above, the more specific use of the Magic Book has yet to be investigated. Since the students in the Type B condition had no restrictions in this regard, they were free to look at pages with solved or unsolved missions, to quickly glance at achieved tokens or to hover them and repeat information, all according to their own liking. However, to accomplish the intended metacognitive effect, the visualization tool most certainly had to be used in a proper manor. Consequently, it is of interest to explore what kind of information the participants attended to.

When studying the distribution of the time spent in the Magic Book on different missions, it seems that this time mainly was dedicated to the first one (see Figure 20). As this graph also reveals, the number of active players steadily decreased with the number of solved missions, and after the third mission, the amount of data is heavily reduced. As expected, students with lower reading performance also progressed more slowly and solved fewer missions, why the calculations for mission 4 to 8 mainly contains data from medium- and high performers.

<sup>3</sup> The possibility of choosing an alternative homework was originally designed for the concurrent ETG-study by

Agata Janiszewska. Used with permission in the study at hand.

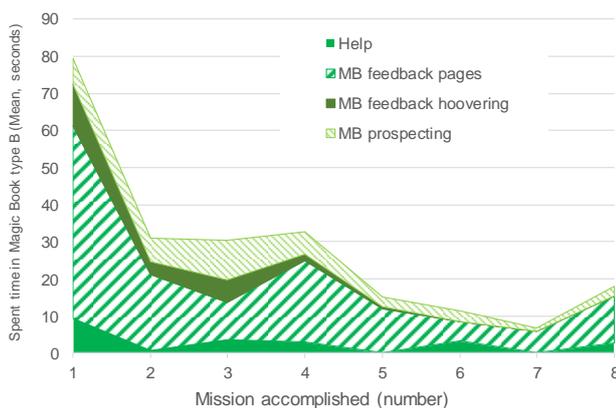


**Figure 20:** Mean values of spent time in the Magic Book per condition and solved mission. Etiquettes show the number of students solving each mission.

When it comes to what type of information the participants attended to, this can be divided into four categories:

- 1) Use of the help-function (same as for Type A)
- 2) Looking at pages and missions with gathered artefacts or at timeline with gathered artefacts, getting feedback and information about learning progress.
- 3) Hovering on gathered artefacts and thereby paying attention to displayed information (text).
- 4) Looking at pages with unsolved tasks and missions or hovering on not yet collected artefacts. This could be regarded as a sort of prospecting, seeking for information about upcoming events or just investigating the tool in general.

The time spent within these four categories is presented in Figure 21 below. As shown, the main part is dedicated to looking at pages, while prospecting or hovering on artefacts to repeat information are less common activities.



**Figure 21:** Attention towards different pages and artefacts in Type B of the Magic Book, per solved mission.

Since reading performance influences metacognition, and since the multiple regression analysis indicates a possible negative effect of the Magic Book on metacognitive monitoring, it also ought to be of interest to analyse what kind of students spending time on the content of the visualization tool. As Table 9 reveals, the differences between performing levels are

quite small, except for prospecting. Students with lower reading capacity seem to spend less time on this activity than others.

**Table 9:** Time spent in the Magic Book per level of reading performance

	Spent time in Magic Book in seconds (M, SD)				
	Reading perf.	Help	MB feedback pages	MB feedback hovering	MB prospecting
Magic Book only Help (type A)	Low	14 (23)			
Magic Book Souvenirs (type B)	Low	24 (25)	99 (58)	19 (43)	16 (19)
Magic Book only Help (type A)	Medium	17 (29)			
Magic Book Souvenirs (type B)	Medium	17 (29)	94 (77)	21 (51)	36 (36)
Magic Book only Help (type A)	High	20 (20)			
Magic Book Souvenirs (type B)	High	16 (26)	97 (103)	28 (38)	38 (42)

Finally, the Magic Book Type B contained a designated area for writing (see Figure 11). To estimate any effect this function could have had on the students' preference for taking notes - as this is regarded as a learning strategy linked to self-regulation and cognitive monitoring - the use of the notation area has been evaluated. Table 10 below shows that most of the students in both conditions did not take notes at all (although encouraged to do so), and no important differences between digital or paper-based notes were discovered.

**Table 10:** Number of students making notes on paper or in Magic Book

	Notes on paper				Notes in Magic Book	
	Nothing	Some	Plenty	Unknown	Nothing	Some
Magic Book only Help (type A)	44	9	5	3		
Magic Book Souvenirs (type B)	42	5	3	1	46	5

#### Additional findings 2: Exploring user preferences

Evaluating behavioural patterns and user preferences by analysing data logs has strong limitations. First of all, it is impossible to be sure of that the actual interactive patterns correlate with certain cognitive processes or conscious choices. Secondly, the logs do not reveal anything about the underlying reasons for their appearance. Thus, to be able to identify possible user needs for metacognitive support in instructional software, the students' own opinions and ideas have been utilized through qualitative data exploration.

In total, 65% of the students, from all performance levels, filled in the questionnaire on user experience (see Appendix C). These responses have, together with the outcomes from group discussions, been evaluated through a thematic inductive analysis (Patton, 2002). The main purpose has been to acquire support, explanations or contradictions to the statistical and numerical outcomes presented above. Herewith, four important aspects of instructional games have been identified. These can all be related to metacognition and learning:

- 1) Challenge
- 2) Simplification or scaffolding
- 3) Exploration
- 4) Learning and meaning

Several students, from all performance levels and conditions, mentioned a wish for **challenge** and competition, e. g. by contesting towards a threatening or mysterious character or by having parallel quizzes and games. Moderate challenges

are known to increase student effort and improve self-regulation (see for instance Clifford, 1990), and these could of course be used as stimulating and engaging ingredients in instructional software. However, challenges are not always cognitive, they can equally be emotional, involve risk-taking or have a mainly visual impact. For a positive impact on metacognitive monitoring, challenging elements probably need to be very carefully designed.

Another relevant finding was the request for *simplification or scaffolding*, i. e. to escape reading a lot of text, getting help from a travelling companion or to use collectables as aids in tasks. A visualization tool could of course be used for easy access to learnt topics or to get an overview of what is to be done (which was the purpose in this study). Several students though, especially low or medium performers, mainly asked for cognitive off-loading by simplifying tasks and content.

An important element in a game with narrative content is the possibility of *exploring* fictive and rich worlds. Medium- and high-performing students from both conditions asked for more and deeper travels, parallel universes or secret paths. Unexpected scenes and mysterious undertones seems to be expected, at least to some extent, in a game like GoH. A tool with tokens could of course be used to unlock such events, or to create cunning meta-tasks in new settings. The present tool did no such thing, although an alternative design is presented under the discussion below.

Finally, the aspect of *learning* was emphasised by several students, as well as the need for *meaningful and logic activities*. Medium- and high-performers made comments on using the archive for storing books and historical inventions from time-travels, or to use it as a kind of library with additional historical information. They also mentioned the possibility of repeating tasks and topics to improve learning. Interestingly, these students mainly came from the baseline condition – without the visualization tool. It is here in its place to mention that in the first mission, a diary from one of the time-travels was placed in the cellar archive as a clickable element, even if temporarily. The view of the archive also displayed rows of bookshelves, although none of these were interactive (see Figure 8).

Single medium- and high-performers, from both conditions, also requested better proofs of achievement, while other mentioned that collectables and tokens should be there for a purpose. The students here seemed quite particular about the importance of meaningful activities in the game. Why gathering objects if these were not to be used? Notably, this request can to some extent be linked to the aspect of simplification and scaffolding (presented above). To use earlier achievements as beneficial tools and use them as leverage for further success is a common element in games (Linderoth, 2012). However, this maneuver is not really compatible with metacognitive monitoring and theoretical education (ibid.).

### Summary

In short, two of three hypotheses had to be rejected. However, the first hypothesis, stating that the visualization tool should

be attended to, was positively confirmed. The participants in the Type B condition (with the full Magic Book) interacted with and attended to the tool significantly more than the participants in the Type A condition. This was evaluated by comparing data logs and measuring the spent time in the Magic Book (in seconds) for each participant.

The second hypothesis, stating that the tool would have a positive impact on metacognitive monitoring, was not confirmed. Minor significant differences in KMA-indexes were found between conditions, but a multiple regression analysis only revealed the variable *reading performance* as positively correlated with metacognitive monitoring. As opposed to the stated hypothesis, the Type B condition resulted in a small but significant *negative* effect on metacognitive monitoring. The time spent in the Magic Book did not affect calculated indexes at all.

The third and last hypothesis, stating that the tool would improve self-regulation, also had to be rejected. The effect of the tool on engagement and endurance was evaluated by calculating and comparing the participants' playtime at home. Spreading from 0 minutes to two hours, the dedicated time to the homework had a large variation, also within groups. No significant differences between conditions were to be found.

To further explore possible reasons for the outcomes presented above, user patterns and user experiences were evaluated. Exploration of data-logs here revealed that the participants did not use the artefact in an optimal way, since they merely scanned the content and did not interact with the gathered souvenirs or read about them. In spite of this, verbal reports from the students (through inquires or discussions) indicated a request for collectables, tokens and proofs of achievement. Some of the students here wanted to be able to repeat information, learn more or improve results. Others merely requested rewards or to get advantages in following missions and tasks.

## 5 Discussion

In summation, the results of the study show that the visualization of acquired knowledge - in the shape of informative tokens in a "Magic Book" - caught the users' attention during play, although it did not seem to have any effect on the students' engagement or self-regulation. As opposed to the stated hypothesis, even though the digital tool was perceived and used, the outcomes imply a minor *negative* impact on metacognitive monitoring and the students' own judgement of what actually is learnt or not. Exploration of data-logs moreover demonstrate that the tool mainly was of interest for the player in the beginning of the game, and that the tokens merely were scanned or briefly observed. The players rarely tried to access the text attached to the tokens, and the digital notebook was only used by a handful of the participants. Finally, additional findings point out the players' wishes for meaningful activities and collectables with historical content. Interestingly, the possibility of saving souvenirs and books from performed time-travels was primarily asked for by students without the visualization tool (within the base-line condition, Type

A). This in turn indicate a metacognitive weakness for requesting functions and effects that perhaps not, in reality, will be used at all during actual play.

The presented results could be due to a number of reasons, not only to the underlying cognitive and metacognitive structures of the human brain or to the learning processes that grow out of them. First of all, since the experimental setup has been both complex, exploratory and partly naturalistic, the method and designed material ought to be discussed and reflected upon. Secondly, the results could depend on aspects of two-dimensional and screen-based digital tools that is hard to translate from other learning material used in a much wider, concrete and multidimensional reality. This subject is also of interest to debate. And finally, as the reader probably already have guessed between the lines, the study mainly ought to be seen as a first attempt to evaluate aspects of knowledge visualization in instructional software. Consequently, the need for further research requires to be ventilated. All of these aspects are discussed in the passages below, always with the distributed cognitive system (presented in Figure 1) in mind.

### *Methodologic considerations*

The chosen methodology primarily rests on experiences drawn from empirical research on educational technology and meta-cognition. These studies have served as guidance and inspiration but are not fully comparable to the present study, why this could be regarded as an unpredictable xenogamy between these research areas. Measures and methods were here combined and effectuated in a new and untested fashion, giving a significant part of the experimental framework an explorative nature.

One important question to ask in this case is if the KMA-test was adequate to use for testing hypothesis no 2 (that is, if metacognitive monitoring was affected by the visualization tool), and if this test was performed in an acceptable manor. Since the historical content in the game was quite specific, and since the playtime in the present study was limited, the knowledge that was manifested in the test could hardly be defined as “general”. As mentioned earlier, the KMA-test is mainly used for evaluating prior knowledge that has been acquired during a longer period and therefore lies deeper in the student’s mind, and the appropriateness of this method within the study at hand could be questioned. Still, this method has strong advantages, not at least in terms of its straight-forwardness. The metacognitive effort for answering quick FOK-questions has to be considered as considerably low in relation to responding to more open and unravelling ones (Tobias & Everson, 2002; Tobias & Everson, 2009).

Of course, the specific questions in both parts of the KMA-test could inflict on the result, as well as the fact that not all of the questions in the latter part was of multiple-choice character (this is normal procedure for KMA-tests, *ibid.*). In addition, to ensure that the outcome reflected the content of the game and nothing else, the questions could have been formulated differently and perhaps more concisely. It is not impossible that

some of the questions in the knowledge-test reflected historical knowledge that the students have accessed outside the *GoH* (as some of them mentioned, orally during the test or in writing). The test could equally have been designed with more weight on the first missions, solved by the majority of the students. As it now where, the calculated Hamann-coefficient often contained more negative (no-incorrect) than positive (yes-correct) answers. Since the coefficient equates these types of responses, it might have evaluated the students’ knowing-of-not-knowing instead of their knowing-of-knowing (see Table 5 for the coefficient’s equation). When it comes to procedural matters, the somewhat unconventional manual and paper-based format worked very well. Although the exact response-time for each FOK-question might have varied slightly between classes, these minor flaws are considered to be of negligible importance for the results at hand.

Additional aspects with effect on the result is the specific design of the stimuli, the longevity of the study and the limitations that originated from the setup with two concurrent studies performed simultaneously. The fact that the use of the visualization tool resulted in a minor negative impact on metacognitive monitoring might be due to the tool’s appearance or functionality. For instance, the blinking of the Magic Book could have caught the user’s attention in a quite opposite way than expected. Perhaps this signal disturbed or annoyed the player instead of informing him or her about important updates?

It is also possible that this kind of digital artefact must be used much more efficiently, within a longer period of time, perhaps in several domains or in a more general way, to have a more significant and positive impact on metacognitive processes and self-regulation strategies. For students to understand and appreciate this kind of tools, they might equally be in need for more direct methodological instructions, repeated testing or situations where they can compare learning and progression with and without it. It is not unlikely that a better awareness of the purpose of the tool would have had a quite different effect on the outcome, not least if the students would have received an opportunity to self-evaluate the benefit of using it.

To evaluate such matters, discussions between students on their experiences and learning strategies would be of great interest. The study would also have benefited from experimental sessions in smaller groups, facilitating user observation and think-aloud protocols while playing. Additional inquiries regarding the use of the Magic Book also could have shed light on the reasons behind the specific use of it, making the evaluation of the logs less speculative. Still, as the present study was performed in cooperation with another ETG-project, the experimental setup had to fit them both, leading to compromises and limitations. The advantage of performing the studies simultaneously was above all practical, both in regard to the acquisition of test persons and to the realizations of the sessions on-site, at the schools. The drawbacks were mainly that the students could only be submitted to a limited number of tests and questions, that the knowledge-test had to fit both

projects and that the material had to contain all possible variants of stimuli.

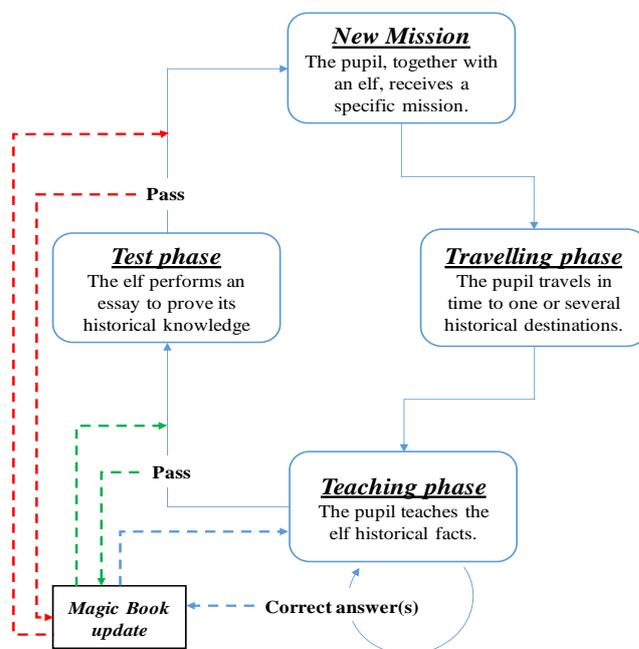
As the theoretical framework is based on a distributed view of human cognition and learning, the gathered data is considered to represent different levels of information within the cognitive system. Consequently, the study would perhaps also have benefited from a slightly different methodological mix in this regard, not at least to get a good balance between data concerning inner mental states, observable behaviour and sociocultural aspects. As always, qualitative data here have to be collected and evaluated with caution, as the design of the game can lead to favourable answers due to priming effects. For instance, in the present case the students didn't suggest the possibility of using a fellow travelling companion as an external memory for performed missions, while several of them responded that the cellar archive could be appropriate for collecting items and books from the time travels. This response might very well be due to the interactive possibility of opening one diary in the beginning of the game, instead of reflecting the players' true wishes and needs for visualized knowledge. Future research ought to bear all these aspects in mind, which is discussed further down in the chapter at hand.

### The nature of digital worlds

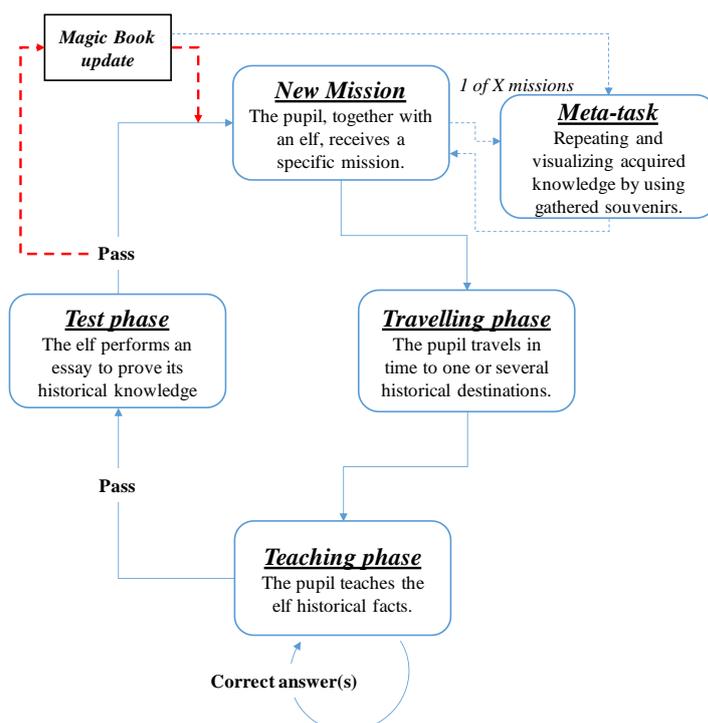
The main purpose of the study at hand has been to verify the need for visualizing acquired knowledge in instructional software. Even though thoroughly designed and experimentally evaluated, the use of a game-related visualization tool did in this case not seem to have any positive impact on metacognitive monitoring. This might, of course, depend on the design and appearance of the artefact, but it could also, at least to some extent, be due to the specific affordances of digital medias in general. As stated earlier, a screen-based game has a sequential nature with limited physical space and dimensionality. This also means that one game-related activity easily overshadows another, simply by visually covering earlier activities with new information. For instance, taking glances at surrounding material and information in the classroom while reading or calculating could be quite easy without losing track of what you are doing. However, displaying such information digitally in instructional software might lead to that the ongoing activity becomes – partly or entirely – hidden, leading to interruptions and disturbances and thereby effecting the cognitive load as well as the user experience.

In fact, the task of designing multileveled digital environments with minimal effect on working-memory is a great challenge and necessitates profound knowledge about human-computer interaction (Benyon, 2010; Gulliksen & Göransson, 2002). Since human beings tend to off-load their memory on the environment by the use of physical triggers and placeholders, parallel tasks and disturbances in the real world often become simplified and less demanding (Dix et al, 2004; Kirsh, 2001). If embedding all pedagogical properties inside an instructional software, overlapping tasks and multiple layers of information should probably be avoided. As metacognition also seem to be intimately related to working-memory (Harris,

Graham, Brindle & Sandmel, 2009), this is perhaps even more important. Another way of raising metacognitive awareness when playing a game like GoH would then be to use the narrative of the game for repeating and visualizing learning progress. This might be done by adding game-elements where new assignments reflect earlier achievements, by forcing the players to attend to their performance through special tasks, or by reformulating dialogues and displayed information. The figures below show the present and alternative conceptual design of game elements with metacognitive content.



**Figure 22:** The present GoH-design, Type B of the Magic Book. The book is updated several times, leading to parallel tasks and attention away from the original game play at three occasions.



**Figure 23:** An alternative GoH-design, with less updates of the Magic Book and without parallel tasks during missions.

On the other hand, instead of trying to assimilate digital artefacts to contain identical or similar properties as non-digital ones, the focus should perhaps lie on the use of a combination of both, utilizing the specific strengths of each one of them. As one of the teachers in one of the participating classes pointed out, a great advantage with instructional software is its linearity. That is, to progress in the game you must do all tasks, you cannot skip a step or jump from one thing to another according to your own liking. This ought to produce a unique opportunity of teaching proper learning strategies, such as taking notes, underlining key words or repeat information. A number of obligatory steps could here be embedded in the software, forcing and encouraging the player to apply a specific beneficial behaviour.

When discussing learning visualization and metacognition in school, the role of the teacher is of course also of the greatest importance (Zohar, 1999; Hattie & Timperley, 2007). It is mainly the teacher that can start discussions, ask meta-questions or display useful information about the students' achievements. Even if this aspect is out of the boundaries of this thesis, the used learning material (digital or not) could facilitate or complicate such matters. Making the learning progress visual and accessible (as in this case, through the opening of a Magic Book with tokens) ought to make it easy for peers to discuss and share their experiences (such as their latest time-travel, a specific destination etc). The tokens would here serve as clues and external memories, and the players could quickly get a summary of the most important aspects of their performances.

It would of course also be possible to use the interface and its content for more playful activities. Perhaps the souvenirs could be used as cards in a game of memory or as inspiration for a fictive story? Such off-task interactions should not only be seen as disturbances or meaningless fun. On the contrary, to let the students rest from demanding cognitive activities by (now and then) engaging in joyful non-task relevant activities can increase engagement and receptivity in the classroom (Gulz, Silvervarg & Sjöden, 2015).

#### *Future research*

Even though the study at hand hasn't resulted in significant measures verifying the importance of visualizing learning in instructional software, the experiences drawn from these experiments are both useful and inspiring. First of all, the results point out that the specific use of digital artefacts is hard to fully predict, and that interactive software with voluntary elements ought to be thoroughly tested before taken into experimental settings. Another important finding is the difficulty of interpreting data logs without supporting inquiries or user observations. Performing research with a distributed view of cognitive states and processes in a naturalistic and somewhat unpredictable setting also has to be considered as challenging and complex, why such studies have to be designed and evaluated with great care. Nevertheless, as Schwartz and Martin (2006) so wisely pointed out, if this theoretical standpoint is to be de-

fended and used within the educational field, studies that verify the impact of artefacts on learning processes have to be carried out.

So, what lies ahead in this area? What sort of questions ought to be asked and how could these be tested properly? In more general terms, any study evaluating newly created interactive artefacts would probably benefit from an experimental design in several steps, where the testing starts within smaller designated groups (still in two conditions) and with an emphasis on qualitative data. After a couple of iterations and design changes, a larger study could be effectuated and quantitatively evaluated. Such a procedure would embrace a more inductive design and perhaps also an element of grounded theory - where the focus lies on generating theory out of data rather than out of a particular theoretical content (see for instance Patton, 2002). Since the research on metacognition and digital learning at present lack specific standards and guidelines, this is perhaps necessary, not at least to draw conclusions from experiences and establish solid and experimentally valid routines for future inquiries in the domain.

When it comes to digital artefacts and their effect on metacognition and learning, a number of interesting aspects could be investigated. For instance, are there any differences between receiving tokens for acquired knowledge and gathering them yourself during play? Research on learning strategies indicate that the very construction of external representations (such as drawing diagrams or symbols) could lead to deeper learning and better transfer than simply perceiving or interpreting them (Ainsworth, 2006; Cox, 1999; Easterday, Alevén & Scheines, 2007). Consequently, it would be of great interest to see if similar effects also could be found when studying digital learning material, and if the very handling of the artefacts also inflicts on metacognitive monitoring. Other design issues is how immediate and specific the feedback in terms of visualized knowledge has to be, if it should be possible to ignore or obligatory to attend to, and if students from different performance levels would be annoyed or disturbed by different kinds of information.

The possibility of using knowledge-based visualization tools (in line with the presented Magic Book) as a transparent GUI between teacher and student or between peers also would be interesting to evaluate. Could this kind of artefact be used for raising metacognitive awareness through open discussions and feedback? In this case, how should it be designed and when would it be appropriate to use it?

As the passage above also has pointed out, the nature of digital worlds has to be taken into consideration in this type of studies. For instance, the strong sequential nature in instructional software has its advantages, but one important question to raise is if a constrained methodological training in this kind of environment could create transfer effects into other (more analogue or flexible) settings. According to a distributed and socio-cultural view, true transfer is hard to establish (Greeno et al, 1996). Nevertheless, perhaps metacognition and self-regulation strategies could compensate for the differences in affordances and restraints between the used artefacts? That is, if students are given opportunities to reflect on their used

methods and to evaluate them, perhaps the benefits of the artefacts' affordances (such as the possibility of marking text or taking notes) could overshadow the lack of constraints in a new context? This would most certainly be of interest to study further.

And finally, in a spirit of distributed learning, evaluating artefacts in relation to their metacognitive support would also be intriguing. For instance, what if the differences in transfer detected by Schwartz and Martin (2006) could be explained by the artefacts' influence (or lack of influence) on metacognitive processes? Exploring deeper cognitive structures when evaluating tools and learning material could be one way of building bridges between the different learning theories presented in the beginning of this paper. This would also shed light on a more holistic and dynamic view of cognition and learning in general.

## Acknowledgements

This study has been designed, performed and reported within a period of 20 weeks. This would not have been possible without aid and assistance from my fellow Master-students at Lund University and the encouragement from mentors and professors. I would especially like to thank Kristian Månsson and Erik Anderberg for assistance with implementing my design ideas, Agneta Gulz and Magnus Haake for guidance and valuable insights, and Agata Janiszewska for a fruitful and joyful cooperation throughout the project. Last but not least, I want to express my gratitude towards all participating and engaged students and teachers. Without you, the study wouldn't exist at all. Thank You!

## References

- Abrami, P. C., Venkatesh, V., Meyewe, E. J. & Wade, C. A. (2013). Using Electronic Portfolios to Foster Literacy and Self-Regulated Learning Skills in Elementary Students. *Journal of Educational Psychology*, *105*, 1188–1209
- Ackerman, M. S, Dachtera, J., Pipek, V. & Wulf, V. (2013). Sharing knowledge and expertise: The CSCW view of knowledge management. *Computer Supported Cooperative Work: CSCW: An International Journal*, *22*, 522-573.
- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *6*, 183-198.
- Area of Figures (2014). <https://itunes.apple.com/se/app/area-of-figures/id434220361?mt=8>
- Arievitch, I. M. & Haenen, J. P. (2005) Connecting Sociocultural Theory and Educational Practice: Galperin's Approach. *Educational Psychologist*, *40*, 155-165.
- Avey, J. B., Aviolo, B. J., Crossley, C. D. & Luthans, F. (2009). Psychological Ownership: Theoretical Extensions, Measurement and Relation to Work Outcomes. *Journal of Organizational Behavior* *30*, 73–191
- Axelsson, A., Andersson, R., & Gulz, A. (2016). Scaffolding executive function capabilities via play-&-learn software for preschoolers. *Journal of Educational Psychology*, *108*, 969-981.
- Benyon, D. (2010). *Designing interactive systems: A comprehensive guide to HCI and interaction design*. (2:d edition) Harlow, England: Addison Wesley.
- Blair, K., Schwartz, D.L., Biswas, G., & Leelawong, K. (2007). Pedagogical agents for learning by teaching: Teachable Agents. *Educational Technology*, *47*, 56-61.
- Brown, J., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning, *Educational Researcher*, *18*, 32-42.
- Butler, P. (2006). *A Review of The Literature on Portfolios And Electronic Portfolios*. Publication for the eCDF ePortfolio Project, Auckland, New Zealand.
- Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7-19.
- Clifford, M. (1990). Students need challenge, not easy success. *Educational Leadership*, *48*, 22-26.
- Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, *9*, 343–363
- de Léon, D. (2003) *Artefactual intelligence. The development and use of cognitively congenial artefacts*. Doctoral Thesis. Lund University.
- Dix, A., Ramduny-Ellis, D. & Wilkinson, J. (2004). Trigger Analysis - understanding broken tasks. I D. Diaper & N. Stanton (red.) *The Handbook of Task Analysis for Human-Computer Interaction*, (pp. 381-400). Lawrence Erlbaum Associates
- Doering, A. & Veletsianos, G. (2009). Teaching with Instructional Software. In M. D. Roblyer & A. Doering (Eds.) *Integrating Educational Technology into Teaching*, (pp. 73-108). Upper Saddle River, NJ: Pearson Education
- Dourish, P., Bellotti, V., Mackay, W. & Ma, C. (1993). Information and Context: Lessons from a Study of Two Shared Information Systems. *Computing Systems*, 42–51.
- Easterday, M. W., Alevan, V. & Scheines, R. (2007). 'Tis Better to Construct than to Recieve? The Effects of Diagram Tools on Causal Reasoning. In *Proceeding of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*. (pp. 93-100). IOS Press.
- Greeno, J., Collins, A. & Resnick, L. (1996) Cognition and Learning. in Berliner, D. & Calfee, R. (eds.) *Handbook om Educational Psychology*, (pp. 15-46), Macmillan, New York
- Gulliksen, J. & Göransson, B. (2002). *Användarcentrerad systemdesign*. Lund: Studentlitteratur.
- Gulz, A. & Haake, M. (2015) Att stödja och utforska lärande med hjälp av digitala läromedel. *Antologi i Utbildningsvetenskap* (Red.) A. Persson & R. Johansson (pp. 45-64).
- Gulz, A., Silvervarg, A. & Sjöden, B. (2015). Design for off-task interaction – rethinking pedagogy in technology enhanced learning. In *The 10th IEEE International Conference on Advanced Learning Technologies (ICALT)*, (pp.204-206).

- Gärdenfors, P. (2010). *Lusten att förstå. Om lärandet på människans villkor*. Natur och Kultur, Stockholm.
- Hall, G. (2002). Associative Structures in Pavlovian and Instrumental Conditioning. In Pashler, Hal / Gallistel, Randy (eds.), *Steven's Handbook of Experimental Psychology*. John Wiley & Sons.
- Happi Stavar (2014) <https://itunes.apple.com/se/app/happi-stavar-korsord-f%C3%B6r-barn-by-happi-papi/id464675842?mt=8>
- Harris, K. H., Graham, S., Brindle, M. & Sandmel, K. (2009) Metacognition and Children's Writing. In D. J. Hacker, J. Dunlosky, A. C. Graesser (Eds.), *Handbook of metacognition in education*, (pp 131-153). New York, Routledge
- Hattie, J., Masters, D. & Birch, K. (2016). *Visible Learning into Action. International Case Studies of Impact*. Routledge, NY.
- Hattie, J. & Timperley, H. (2007) The Power of Feedback. *Review of Educational Research*, 77, 81-112.
- Huffaker, D. A. & Calvert, S., L. (2003). The New Science of Learning: Active Learning, Metacognition, and Transfer of Knowledge in E-Learning Applications. *Journal of Educational Computing Research*, 29 (3), 325-334.
- Hutchins, E. (1995). "How a cockpit remembers its speeds". *Cognitive Science*, 19, 265–288.
- Håkansson (2011). *Synligt lärande. Presentation av en studie om vad som påverkar elevers studieresultat*. Sveriges kommuner och landsting. Publikation.
- Johansson, P., Hall, L. & Sikström, S. (2008). From change blindness to choice blindness. *Psychologia*, 51, 142-155.
- Joseph, N. (2009). Metacognition Needed: Teaching Middle and High School Students to Develop Strategic Learning Skills, *Preventing School Failure: Alternative Education for children and Youth*, 54, 99-103,
- Kirsh, D. (1996). Adapting the environment instead of oneself. *Adaptive Behavior*, 4, 415-452.
- Kirsh, D. (2001) The context of work. *Human-Computer Interaction*, 16, 305-322.
- Kirsh, D. & Maglio, R. (1992). Perception action in Tetris. In R. Simmons (chair) AAA spring symposium on control of selective perception. Stanford University.
- Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge: Cambridge University Press.
- Linderoth, J. (2010). *Lärarens återkomst. Från förvirring till upprättelse*. Natur och Kultur, Stockholm.
- Linderoth, J. (2012). Why gamers don't learn more: An ecological approach to games as learning environments. *Journal of Gaming and Virtual Worlds*, 4, 45-61
- Lindström, P., Gulz, A., Haake, M. & Sjöden, B. (2011). Matching and mismatching between the pedagogical design principles of a maths game and the actual practices of play. *Journal of Computer Assisted Learning*, 27, 90-102
- Meyer, E., Abrami, P., Wade, A., Aslan, O. & Deault, L. (2010). Improving literacy and metacognition with electronic portfolios: Teaching and learning with ePEARL. *Computers & Education*, 55, 84-91.
- Ohlsson, S. (2008) Computational models of skill acquisition. In Ron Sun (ed.). *The Cambridge Handbook of Computational Psychology*. Cambridge University Press. (pp. 359–395).
- Paris, S. G. & Paris, A. H. (2001). Classroom Applications of Research on Self-Regulated Learning. *Educational Psychologist*, 36, 89–101
- Patton, M. Q. (2002) *Qualitative Research and Evaluation Methods*. (Third edition) London: Sage Publications Inc.
- Pierce, J. L., Kostova, T. & Dirks, K. T. (2003). The state of psychological ownership: Integrating and extending a century of research. *Review of General Psychology*, 7, 84-107
- Sawyer, R. K. & Greeno, J. (2009). Situativity and Learning. In P. Robbins & M. Aydede (Eds.). *The Cambridge handbook of situated cognition*. Cambridge University Press. (pp. 55-77).
- Schneider, W. (2008). The Development of Metacognitive Knowledge in Children and Adolescents: Major Trends and Implications for Education. *International Mind, Brain, and Education*, 2, 114-121
- Schwartz, D. L. (1999). The productive agency that drives collaborative learning. In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 197-218). NY: Elsevier Science.
- Schwartz, D. & Martin, T (2006) Distributed learning and mutual adaptation. *Pragmatics & Cognition* 14, 313-332.
- Shettleworth, S. J. (2013). *Fundamentals of Comparative Cognition*, Oxford University Press.
- Shute, V. (2007). Focus on Formative Feedback. *Review of Educational Research*, 78, 153–189.
- Sjöden, B. (2015). Vad är ett bra digitalt läromedel? *Antologi i Utbildningsvetenskap* (Red.) A. Persson & R. Johansson. (pp. 79-94).
- Tobias, S. & Everson, H.T. (2002). *Knowing What You Know and What You Don't: Further Research on Metacognitive Knowledge Monitoring*. College Board Rep. No. 2002–03). New York: College Board
- Tobias, S. & Everson, H.T. (2009). The importance of knowing what you know. A Knowledge Monitoring Framework for Studying Metacognition in Education. In Hacker, D.J., Dunlosky, J. & Graesser, A.C. (ed), *Handbook of metacognition in education* (pp. 107-123). New York, Routledge
- Trevitt, C., Macduff, A. & Steed, A. (2014). [e]portfolios for learning and as evidence of achievement: Scoping the academic practice development agenda ahead. *Internet and Higher Education*, 20, 69-78.
- Van Voorhis, W. & Morgan, L. M. (2007). Understanding Power and Rules of Thumb for Determining Sample Sizes. *Tutorials in Quantitative Methods for Psychology*, 3, 43-50.
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Af-flerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1, 3–14.

- Wilson, R. A. & Clark, A. (2009) How to Situate Cognition, The Situated Perspective on Emotion. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 55-77). Cambridge: Cambridge University Press.
- Winne, P. H. & Jamieson-Noel, D. (2002) Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology*, 27, 551–572
- Winne, P. H. & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky, A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 259-277). New York, Routledge
- Winne, P. H., Nesbit, J. C., Kumari, V., Hadwin, A. F., Lajoie, S. P., Azevedo, R. & Perry, N. (2006). Supporting Self-Regulated Learning with gStudy Software: The Learning Kit Project. *Technology, Instruction, Cognition and Learning*, 3, 105-113.
- Wright, D. B. (1996). Measuring Feeling of Knowing: Comment on Schraw (1995). *Applied Cognitive Psychology*, 10, 261-268
- Zimmerman, B. J. & Cleary, T. J. (2009). Motives to Self-Regulate Learning: A social cognitive account. In Wentzel, K. & Wigfield, K. (red), *Handbook of Motivation at School* (pp. 247-264). New York, Routledge
- Zohar, A. (1999). Teachers metacognitive knowledge and the instruction of higher order thinking. *Teaching and Teacher Education*, 15, 413-429



## Appendix B

Questions in Knowledge-test and their relation to FOK-test in appendix A (in Swedish).

The questions elicited from the final index calculations are presented in parentheses.

### FOK-rela-

#### tion

##### 0a. Kryssa för dom uppdrag som du och tidsalven klarat av tillsammans!

Inez  Sofia Brahe  Industrialismen och barns villkor under denna tid   
Pesten  Emelie du Chatelet  Galileo Galilei  Isaac Newton

0a

##### 0b. Kryssa för dom resor som du har genomfört!

Till Eric och hans familj som var på resa genom Sverige  Till domstolen i Stockholm   
Till Inez hemma  Till uppfostringsanstalten  Till Ven  Till Inez i skolan   
Till familjen Horn som hade fått en målning av sitt döda barn  Till Florens i Italien   
Till London  Till bondgården  Till Oxford i England  Till Cirey i Frankrike

0b

##### 0c. Kryssa i vilket århundrade som passar ihop med rätt uppdrag!

	Ej utfört	1400	1500	1600	1700	1800
Sofia Brahe						
Pesten						
Industrialismen och barns villkor						
Inez						
Emelie du Chatelet						
Isaac Newton						
Galileo Galilei						

0c

- (1. Vad är en historisk källa?) (1)
- (2. Skriv ner de fem frågorna man ska ställa när man värderar en källa.) (2)
3. Skriv ner 3 av de olika sorters källor du kommer ihåg från lärspelet. Inte vad källan var utan vad för SLAGS källa. 3
4. Om du vill ta reda på hur det var att leva på 1700-talet, vilken källa eller källor är bäst för att få mest information, och mest användbar information? Ringa in det alternativet du tror är bäst. 4
- a) Flera olika saker som kläder, möbler och hus.  
b) Flera domstolsprotokoll  
c) En dagbok  
d) En dagbok, olika saker som kläder, och domstolsprotokoll. 5
5. Varför hamnade barn på uppfostringsanstalt? Ge två exempel. 6
6. Vad betydde det när allmän skolplikt infördes? Ringa in rätt alternativ. 6
- a) Allmän skolplikt betydde att alla barn i staden var tvungna att gå i skolan, men barn på landet behövde inte  
b) Alla barn var tvungna att gå i skolan  
c) Det betydde att alla barn som ville kunde gå i skolan om de inte ville jobba.
7. När infördes allmän skolplikt i Sverige? Ringa in rätt alternativ. 7
- a) 1682  
b) 1982  
c) 1882
- (8. Varför är det viktigt att ha flera historiska källor av samma händelse?) (8)
- (9. Vad är aga?) (9)
10. Var härjade pesten 1665-1666 enligt Samuel Pepy's dagbok? 10
11. Vad trodde man PÅ DEN TIDEN att pesten spreds av? Ringa in rätt alternativ. 11
- a) Man trodde att pesten spreds av råttor

- b) Man trodde att pesten spreds av en farlig ånga i luften, miasma, och att den spreds av hundar och katter.  
c) Man trodde pesten spreds av var från öppna sår. Man kunde smitta genom kroppskontakt.
12. Vad vet man NU att pesten spreds av? Ringa in rätt alternativ. **12**  
a) Pesten spreds av främst av fåglar, såsom tamhöns och kråkor.  
b) Pesten spreds av katter och hundar  
c) Pesten spreds främst av råttor.
13. Vad studerade Sofia Brahe? Ringa in rätt alternativ. **13**  
a) Hon studerade astronomi och astrologi  
b) Hon studerade gastronomi och arkeologi  
c) Hon studerade bibeln och andra religiösa skrifter  
d) Hon studerade biologi och filateli
14. Vad experimenterade Sofia Brahe med och vad försökte hon göra? Ringa in rätt alternativ. **14**  
a) Experimenterade med explosioner och tillverkade fyrverkerier  
b) Experimenterade med värme och tillverkade ett element  
c) Experimenterade med blommor och försökte skapa nya arter av grönsaker  
d) Experimenterade med att göra guld av andra metaller och tillverkade örtmedicin
15. Vad studerade Galileo Galilei? Ringa in rätt alternativ. **15**  
a) Jordens form  
b) Optiska fenomen och hur solens strålar bryts  
c) Naturfilosofi och solens och planeternas rörelser  
d) Kemi och jordens beståndsdelar
16. Vad gjorde Galilei för experiment och upptäckter? Ringa in rätt alternativ. **16**  
a) Experiment med fallande och rullande klot  
b) Spegel och illusioner  
c) Åkte jorden runt för att upptäcka att jorden var rund och inte platt  
d) Kokpunkten av ämnen för att visa hur jorden blev till
17. Vad studerade Newton? Ringa in rätt alternativ. **17**  
a) Naturkrafter  
b) Gravitation och kroppars rörelser  
c) Äpplen och andra frukter  
d) Acceleration av kroppar
18. Vad gjorde Newton för experiment och upptäckter? Ringa in rätt alternativ. **18**  
a) Experimenterade med ljus och upptäckte användning av speglar i teleskop  
b) Experimenterade med dynamit och upptäckte tidsinställda bomber  
c) Experimenterade med att tillverka guld genom prismor och upptäckte konstgjorda diamanter  
d) Experimenterade med trögflytande vätskor och upptäckte plexiglas
19. Vad studerade Emelie de Chatelet? Ringa in rätt alternativ. **19**  
a) Meteorologi, biologi och dendrologi  
b) Astrologi och horoskop  
c) Naturfilosofi, matematik och språk  
d) Litteratur, lingvistik och samhällskunskap
20. Vad gjorde Chatelet för experiment? Ringa in rätt alternativ. **20**  
a) Mediciner och kemiska preparat  
b) Spridning av värme och fallande klot  
c) Jordens omloppsbana och matematiska formler  
d) Fysik och beteende

## Appendix C

Open ended questions on user experience and game design<sup>4</sup> (in Swedish)

- 1) Vad tyckte du var roligast med spelet?
- 2) Vad var tråkigast? Tycker du att något ska ändras?
- 3) Vad tycker du ska finnas i arkivet? Vad ska man kunna göra där?
- 4) Om du skulle uppfinna en ny figur i spelet, hur skulle den se ut? Vad skulle den göra där?

---

<sup>4</sup> These questions were formulated by Agata Janiszewska to be used for further game development and interaction design. Used with permission in the study at hand.