Using concept mapping to measure changes in interdisciplinary learning during high school
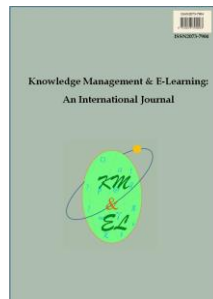
Priit Reiska
Katrin Soika
Tallinn University, Estonia
Alberto J. Cañas
Institute for Human & Machine Cognition (IHMC), USA

Using concept mapping to measure changes in interdisciplinary learning during high school

Priit Reiska*

School of Natural Sciences and Health
Tallinn University, Estonia
E-mail: priit.reiska@tlu.ee

Katrin Soika

School of Natural Sciences and Health
Tallinn University, Estonia
E-mail: katrin.soika@tlu.ee

Alberto J. Cañas

Institute for Human & Machine Cognition (IHMC), USA
E-mail: acanas@ihmc.us

*Corresponding author

Abstract: How, when and what kind of learning takes place are key questions in all educational environments. School graduates are expected to have reached a development level whereby they have, among many fundamental skills, the ability to think critically, to plan their studies and their future, and to integrate knowledge across disciplines. However, it is challenging to develop these skills in schools. Following existing curricula, disciplines are often taught separately and by different teachers, making it difficult for students to connect knowledge studied and learned from one discipline to that of another discipline. The Next Generation Science Standards on teaching and learning natural science in the United States point out important crosscutting concepts in science education (NGSS, 2013). In Estonia, similar trends are leading to an emphasis on the need to further develop scientific literacy skills and interdisciplinary learning in students. The changing environment around us must be reflected in changes in our school system. In this paper, we report on research that intends to answer the questions: (a) "How much do Estonian students develop an interdisciplinary understanding of science throughout their high school education?", and (b) "Is their thinking more interdisciplinary after two years of studies in an Estonian high school?" Additionally, we analyzed the results based on the type of school the students attended, and we examined the use concept mapping to assess interdisciplinary learning. This research is part of an overall study that involved students from 44 Estonian high schools taking a science test similar to the three-dimensional Programme for International Student Assessment (PISA) test (hereafter called PISA-like multidimensional test) as well as constructing concept maps, while in 10th and 12th grade. In this paper, we report on the analysis of the results for 182 of the students, concentrating on the analysis of the concept maps they constructed. The results suggest that there were changes in the students' interdisciplinary knowledge, but these were small and varied depending on the students' school type. They also suggest that changes may be needed in the Estonian educational system to increase the students' level of

interdisciplinary understanding of science.

Biographical notes: Priit Reiska is Vice Rector and Professor for Science Education in Tallinn University, Estonia. His main research interest is using concept mapping in education. In addition, his research interests include relevance of science education, interdisciplinary science education and computer simulations in science education.

Katrin Soika is a doctoral student at the Institute of Educational Sciences in Tallinn University. Her research interests are learning and teaching in natural sciences, impact of animations and different use of concept mapping, including using concept mapping for assessment.

Alberto J. Cañas is Co-Founder, Associate Director and a Senior Research Scientist at the Institute for Human and Machine Cognition – IHMC, where he has led the research and development of the CmapTools concept mapping software. He earned a Bachelor's degree in Computer Engineering from the Instituto Tecnológico de Monterrey, Mexico, and a Master's Degree in Computer Science and a Ph.D. in Management Science, both from the University of Waterloo, Canada. He is interested in the theoretical aspects and in the implementation details of concept mapping in education, uses of computers in education, and knowledge management.

1. Introduction

We assume that there is a general desire that citizens are considerate towards each other, have empathy, are willing and able to collaborate, think critically, act wisely, and are able to connect knowledge from different fields and able to think in an interdisciplinary way. Consequently, these are some of the many behavioral characteristics and competencies which we would like our students to develop (Holbrook & Rannikmäe, 2009; Haridus-JA Teadusministeerium, 2018; National Research Council, 2012). More specifically, we are concerned with students' overall ability to connect knowledge from different fields and their ability to integrate disciplines. Schools, however, tend to be discipline-based and teachers are mainly focused on the topic they must teach, without giving much consideration for the reality of the topic in context of the world around us, and with little intent in engaging with other topics (Henno, 2015). We question whether such a learning environment leads to an interdisciplinary integration of the various topics by students.

This concern led us to the research effort we report in this paper. We intent to measure the level of interdisciplinary understanding of science topics by Estonian high school students and how it evolves as the students advance through school. Our interest is also based on the deep connection between interdisciplinarity and scientific literacy. We believe that interdisciplinarity is one of the highest competences in scientific literacy (Bybee, 1997) and assessing it should give an indication of how able students are in connecting knowledge from different disciplines (Mansilla & Duraisingh, 2007). Assessing interdisciplinarity, however, is a growing concern in the literature, since traditional assessment methods are often not flexible enough or not applicable to measuring interdisciplinarity (Mansilla & Duraisingh, 2007; Stowe & Eder, 2002; Borrego, Newswander, McNair, McGinnis, & Paretti, 2009; Schaal, Bogner, & Girwidz,

2010; Nissani, 1997). Some authors (e.g., Borrego et al., 2009; Schaal et al., 2010) have pointed out the viability of using concept mapping as a tool for assessing interdisciplinarity. Furthermore, concept mapping has been shown to be effective in bringing out the schema of students' pre- and new-learnt knowledge structures (Soika & Reiska, 2014a; Borrego et al., 2009), and is widely used in teaching, learning, planning as wells as for assessment (Borrego et al., 2009; Kinchin, 2011, 2017; Schaal et al., 2010; Novak, 2010; Cañas, Bunch & Reiska, 2010; Cañas, Reiska, & Möllits, 2017; Anohina-Naumeca, 2015). In this research effort, we used concept mapping to assess interdisciplinarity in the students' understanding.

Additionally, we are interested in a large-scale evaluation of the viability of using concept mapping to assess students' depth of understanding of interdisciplinarity, and finding out whether we could carry out the assessment using automatic analysis and evaluation of the students' concept maps. We examined several ways of analyzing concept maps that seemed to suggest changes in the students' knowledge. The concept maps prepared by students were compared to a PISA-like multidimensional test that the students also solved (OECD, 2016; Henno, 2015). The sample of students for the study took into account the variety of schools in Estonia in terms of their results in state-administered exams, and thus included students from schools: a) with very good results on state exams; b) with average results on state exams; and c) with low results on state exams. Each student was presented 30 concepts and a focus question as input for construction of a concept map. A pre-concept map was constructed in 10th grade and a post-concept map in 12th grade. Students were assigned into one of four groups: biology, chemistry, physics and geography. In this paper, we present the results of analyzing the concept maps from the chemistry group.

To assess the concept maps for interdisciplinarity we introduce a numeric Interdisciplinary Quality Index (hereafter called IQI), which was derived from an extensive analysis and evaluation of the set of concept maps. Based on this IQI, we compared students' 10th and 12th grade concept maps to assess not only the quality of the maps but also the degree of interdisciplinarity shown.

The aims of the study were:

1) To investigate how students' interdisciplinarity understanding changes throughout the high school studies.
2) To compare differences and changes in interdisciplinarity understanding among students from different types of Estonian schools.
3) To develop the IQI as a measure of the level of interdisciplinarity understanding expressed in concept maps (Reiska & Soika, 2015; Soika & Reiska, 2014a).
4) To evaluate the feasibility of automatically assessing a large number of concept maps to measure the level of interdisciplinarity understanding expressed by the map builders.

2. Theoretical background

2.1. Meaningful learning

Novak (2010) writes that we acquire new knowledge thru cognitive or meaningful learning, by assimilating and linking new information to our previously acquired knowledge. As a result, students who learn meaningfully can explain newly constructed

knowledge themselves and understand how the newly studied material fits with the knowledge that they already possessed. It is said that through these effective cognitive processes learning is more effective and newly acquired knowledge remains in memory for a longer time period (Klassen, 2006; Novak, 2010). Meaningful learning is based on Ausubel's Assimilation Theory (Ausubel 1968; Novak, 2010), which states that three conditions are required for meaningful learning to take place:

1)   the students should have the relevant prior knowledge;

2)   the learning material should be meaningful; and

3)   the learner should want to learn meaningfully (Bretz, 2001; Novak, 2010; Emenike, Danielson, & Bretz, 2011).

When these three conditions are met, meaningful learning can take place. Meaningful learners tend to have a better organized cognitive structure that facilitates a better understand of daily surrounding processes, and enables them to gain a higher level of scientific literacy (Novak, 2010; Kinchin, 2011, 2017; Cañas, Novak, & Reiska, 2015; Cañas et al., 2017).

2.2. Nature of learning curve

Researchers have found patterns that describe the relationship between learning and experience (Ngwenyama, Guergachi, & McLaren, 2007; Novak, 2010; Klassen, 2006). One of these patterns is referred to as the power of learning, or learning curve. We run into the essence of it in daily processes. Kenneth J. Arrow (1962), one of the first researchers to examine the learning curve, manifested that knowledge increases with time and experiences. Ngwenyama et al. (2007) described learning as a product of experiences which allow an individual to construct knowledge. A learning curve illustrates improvement rates in learning by showing that most tasks are performed faster with practice, and the rates and shapes of improvement are quite similar even for different tasks. (Ritter & Schooler, 2001; Ngwenyama et al., 2007; Adler & Clark, 1997; Benzel & Orr, 2011).

There are different phases within learning curves: (a) an initial steep phase (active learning phase), where the learning occurs and thus the reason for developing faster performance; (b) a plateau phase, where we can expect little improvement in performance (experts are usually in this phase) (Passerotti et al., 2015; Ngwenyama et al., 2007; Ritter & Schooler, 2001). The plateau phase tends to be flat, but there still are small improvements that can be seen after months or even years of practice (Ritter & Schooler, 2001). From a learning perspective, a new learning curve initiates after the end of a previous one, as students begin to study something new (Passerotti et al., 2015; Ngwenyama et al., 2007).

2.3. Scientific literacy

For students to understand connections between concepts and to be able to apply the studied material knowingly, students need to learn meaningfully (Novak, 2010; Cañas et al., 2010; Ruiz-Primo, Schultz, & Shavelson, 1997). In natural sciences, this is referred to as enhancing scientific literacy. From within the various definitions for scientific literacy, we use Holbrook & Rannikmäe's (2009, p. 286) who state that "scientific literacy is an ability to creatively utilize appropriate evidence-based scientific knowledge and skills, with relevance for everyday life and career and solving personally challenging yet meaningfully scientific problems as well as making responsible decisions". There are also

different levels of scientific literacy as brought forth by the Biological Science Curriculum Study (BSCS) (1993) and Bybee (1997) as: (a) nominal literacy (the lowest), (b) functional literacy, (c) structural literacy (or conceptual literacy and procedural literacy) and (d) multidimensional literacy (it is the highest level and students should be able to work independently, link ideas across scientific disciplines, etc.). The last level also denotes that students are expected to possess interdisciplinary knowledge.

2.4. Interdisciplinary understanding

There is no consensus on the definition of interdisciplinary understanding or learning. Many of us understand the term, but do we know what it means? In this study, we use Mansilla and Duraisingh's (2007) definition (p. 219): "We define interdisciplinary understanding as the capacity to integrate knowledge and modes of thinking in two or more disciplines or established areas of expertise to produce a cognitive advancement." Almost the same definition is used by Nissani (1997). Ivanitskaya, Clark, Montgomery, and Primeau (2002) state that interdisciplinary learning needs to create more holistic knowledge than disciplinary learning and interdisciplinary knowledge leads to a complex and internalized organization of knowledge.

Just as there are difficulties in defining the concept, there are also difficulties in assessing the outcome of interdisciplinary learning. In their literature review, Mansilla and Duraisingh (2007) concluded that the literature converges on some premises: (a) an assessment tasks should invite students to build and demonstrate mastery of "whole" performances; (b) criteria and standards should be shared between faculty and students; and (c) assessment should be ongoing and should provide feedback to support learning. Schaal et al. (2010) in their literature review recognized that while interdisciplinary learning needs to be assessed, traditional tests often flunk at its assessment and recommend using concept mapping instead. You, Marshall, and Delgado (2018) comment that they started working on assessment for interdisciplinary learning because of the lack of instruments available.

2.5. Assessment

There are many different ways of assessing knowledge, and the instructor, the students, and the researcher or tutor need to be able to choose the best and most appropriate assessment instrument (Klassen, 2006; Novak, 2010; Stowe & Eder, 2002). Stowe and Eder (2002, p. 80) quote Angelo's (1995): "Assessment is a means for focusing our collective attention, examining our assumptions, and creating a shared culture dedicated to continuously improving the quality of higher learning. Assessment requires making expectations and standards for quality explicit and public; systematically gathering evidence on how well performance matches those expectations and standards; analyzing and interpreting the evidence; and using the resulting information to document, explain, and improve performance."

Assessment should be ongoing and support the student's development, not only controlling learnt truths for grading (Novak, 2010; Stowe & Eder, 2002).

2.6. Concept mapping and assessment with concept maps

2.6.1. Nature of concept mapping

Concept maps, developed by J. Novak and his research team in the 1970's (Novak & Gowin, 1984), are widely used in education as a tool for teaching, studying, learning and assessment, and is based on Ausubel's (1968) theory of meaningful learning. A concept map built by a student to express his or her understanding about a topic is meant to be an external representation of the meaningful connections that are made as new concepts are integrated with previous knowledge in the student's cognitive structure. Thus, a concept map has the form of a hierarchical network of concepts represented as nodes, and linked through the meaningful connection of concepts expressed as linking phrases. Every two concepts are connected together through a linking phrase to form a binding expression called a proposition. This network is expected to reflect a student's (or group of students') personal understandings and misunderstandings, it represents the student's cognitive structure. (Novak, 2010; Kinchin, 2011, 2015, 2017; Ruiz-Primo et al.,1997; Cañas et al., 2015; Schwendimann, 2014; Tao, 2015; Cañas et al., 2017).

Concept maps can be drawn using pencil and paper, but computer software facilitates the construction and revision of the maps in the same way that word processors facilitate writing. Teachers and researchers can assess the digital concept maps through automated tools, allowing them to compare and find concepts or misconceptions (Schaal et al., 2010; Cañas et al., 2010; Soika & Reiska, 2014a; Anohina-Naumeca, 2015; Tao, 2015; Miller, 2008; Novak & Gowin, 1984). There have been questions about the need for training before students are able to construct reasonable concept maps, but one of our previous efforts (Soika & Reiska, 2013) reports that students like to create concept maps using computers and their results do not depend on computer handling skills nor on their previous experience creating computer-based concept maps. Additionally, Schaal and his team (2010) agreed that constructing concept maps on-screen is effective and intuitive for learners. There are a variety of different computer programs and environments to choose from for constructing concept maps using a computer (Cañas et al., 2015; Anohina-Naumeca, 2015; Kinchin, 2015; Tao, 2015). For this research, we used the IHMC CmapTools software toolkit (Cañas et al., 2004; Cañas et al., 2010; Cañas et al., 2015).

Research has shown that by modifying the instructions and input that are provided to students, such as providing the focus question, a list of concepts, or even a skeleton concept map, different situations can be put together that affect the resulting concept maps (Miller, 2008; Cañas, et al., 2015). Thus, it is important to note that we need to be very careful when comparing concept maps that have been constructed under different conditions, since differences in the concept maps may reflect differences in conditions, instructions, input, students' feelings, etc. (Reiska & Soika, 2015; Soika & Reiska, 2014a, 2014b, 2014c; Anohina-Naumeca, 2015; Ruiz-Primo, Schultz, Li, & Shavelson, 2001).

2.6.2. Assessing with concept mapping

There are dissensions in the literature on using concept mapping as an assessment tool (Borrego et al., 2009; Ruiz-Primo & Shavelson, 1996; Ruiz-Primo et al., 1997; Ruiz-Primo, 2004; Anohina-Naumeca, Grundspenkis, & Strautmane, 2011; Anohina-Naumeca, 2015; Miller, 2008; Cañas, et al., 2015; Kinchin, 2011; Schaal et al., 2010; Schwendimann, 2014; Tao, 2015). As the emphasis in schools to provide a more meaningful education strengthens, there is a greater need for more flexible assessment methods, and concept mapping as an assessment tool is one such method. There have

always been controversial statements on the nature of assessment (Klassen, 2006; Borrego et al., 2009; Stowe & Eder, 2002), but researchers occasionally agree that we do need better assessment tools, in particular tools that support learning. Anohina-Naumeca (2015) suggests concept mapping is generally used for summative assessment, but it can also be used as a formative assessment tool. There have been many discussions on how to assess concept maps, e.g. whether to split the concept map or to observe the structure in its entirety. But evaluating the structure is not enough. E.g., Austin and Shore (1995) pointed out that a higher number of links do not guarantee a better understanding of the topic by the student, as many links can be invalid or trivial. We need to also, and mainly, assess the content of the concept maps. Furthermore, for describing semantic changes between concepts, there is a need for measuring the quality of the map (Kinchin, 2011; Schwendimann, 2014). Cañas, et al. wrote in 2015 (p. 9): "… because of the nature of the work, the evaluation of the quality of maps in other applications is not done in as formal a way as in education." The same opinion is pointed out by Borrego and her research team (Borrego et al., 2009).

We decided to use concept mapping as a research instrument for our work after some of our previous efforts pointed out that concept maps could show interdisciplinary understanding by the students in a way that was hard to determine with usual testing methods. Additionally, the literature pointed out the need for more research on interdisciplinarity research with concept mapping (Borrego et al., 2009; Schaal et al., 2010).

The difficulties in assessing scientific literacy and interdisciplinary learning are partly due to the lack of flexible assessment tools. Some authors (Borrego et al., 2009; Schaal et al., 2010; Soika & Reiska, 2014a, 2014b) have found that concept mapping can be useful in assessing these competences. Although there are many methods discussed in the literature for assessing concept maps, some authors suggest that further research is needed (Schaal, et al., 2010; Borrego et al., 2009). There are many ways to evaluate a concept map. For example, it is possible to assess concept maps by comparing them with a map built by an expert (Miller, 2008; Ruiz-Primo, 2004; Tao, 2015); or analyzing the concept maps' structure by counting its hierarchy levels, the number of propositions and branch points, number of orphan concepts, calculate its topological taxonomy score, calculate values by different rubrics, and many other forms reported in the literature (Novak, 2010; Kinchin, 2015; Schaal et al., 2010; Cañas et al., 2010; Soika & Reiska, 2014a). It should be noted that quantity measures, although easy to calculate, do not represent the content expressed in the concept map (Kinchin, 2011). An assessment of the content, in terms of quality of propositions, response to the focus question, and overall quality of the map must also be undertaken (Reiska & Soika 2015; Soika & Reiska, 2014a; Borrego et al., 2009; Cañas et al., 2015).

Cañas et al. (2015) write that a good concept map has a good graphical structure and content, and additionally a good overall map quality. They further suggest that a good concept map responds to the focus question, but an excellent concept map explains the problem in a clear fashion.

2.6.3. Computer based analysis and concept mapping

A thorough analysis of a concept map, in particular when it involves not only evaluating the structure but also the quality of the content of the map, takes time. The examination of a large number of concept maps is thus simplified by using software tools, even with the inconvenience that they tend to emphasize quantitative rubrics (Anohina-Naumeca,

2015; Cañas et al., 2010; Tao, 2015). For this study, we use the CmapAnalysis program (Cañas et al., 2010), which generates results that can be further manipulated in MS Excel and allows both quantitative and qualitative measures of concept maps. For this study, the main measures that were evaluated were:

a) Proposition count: the number of propositions ("sentences") in the concept map;

b) Branch points: the total number of concepts and linking phrases that have at least three connections;

c) Propositions with a score of 2: count of propositions that were assessed as correct and well-explained sentences;

d) Discipline-based or intra-cluster proposition count: propositions (sentences) that were created from concepts from the same cluster (discipline);

e) Inter-cluster proposition count: propositions (sentences) that were created from concepts between different clusters (disciplines);

f) Central concept: concept that has the highest number of propositions (sentences) linked to and from it (largest branching point) (Cañas et al., 2015; Soika & Reiska, 2014a).

2.6.4. Assessing interdisciplinarity with concept mapping

There are few studies where interdisciplinarity is identified using a concept maps. One such study was carried out by Borrego and her team in 2009. The study included pre- and post-concept maps with 11 students and claims (Borrego et al., 2009, pp 22): "Concept maps, as we have shown here, are robust tools for evaluating knowledge integration in interdisciplinary settings, particularly, as described above, when the process of selecting and training scorers takes disciplinary differences into account." And on page 21 they write "... given the centrality of knowledge integration in interdisciplinary environments and the power of concept maps to represent complex knowledge networks, we argue here that concepts maps are a valuable tool for assessing students' interdisciplinary development." Their study's rubric considers three different measures: comprehensiveness (covering completely/broadly), organization (to arrange by systematic planning and united effort) and correctness (conforming to or agree with fact, logic, or known truth). They had different experts manually mark and assess the concept maps and the results show that concept mapping can be used for assessing, but the manual process is time consuming and gives rise to differences in opinion (and discussions) among experts. We intend to circumvent these issues with the use of computer-based concept map assessment tools.

3.   Methodology

3.1. Structure of the research

In this chapter we describe the main study, as shown in Fig. 1, for which data collection lasted 3 years. Previously pilot studies were done (Soika & Reiska, 2013; Soika & Reiska 2014a, 2014b, 2014c) that provided input for the valid and reliable research instrument we designed. In the present study high school students (in the 10th and 12th grade) were asked to construct pre- and post-concept maps and completed pre- and post-PISA-like tests. In this paper we compare the results of the PISA-like test and the concept maps

created by the same students. That gave us the opportunity for a more in-depth investigation of concept mapping.

3.2. Sample of the study

This study is based on a subset of the data from the large-scale natural science scenario-based longitudinal study Lotegym (Soobard, Rannikmäe, & Reiska, 2015; Laius, Post, & Rannikmäe, 2016) which was carried out in 2011-2014 (illustrated in Table 1). There were N_1=1614 students (ages 16-19) from different Estonian high schools. Students were examined with a PISA-like multidimensional test and concept mapping. The goal of the test was to investigate the scientific literacy level of high school students. There were differently designed parts of the exercise that controlled students' skills. Parts were designed by the SOLO taxonomy. (Soobard et al., 2015). Exercises were presented as multiple choice and open-ended questions. Students had to solve a scientific problem, make a decision, and choose a correct scientific explanation during the exercise. Results of the exercises were coded on a three points scale (Soobard et al., 2015; Laius et al., 2016). The exercises were based on different fields and scenario-based topics in natural science and the themes were from biology, chemistry, physics and geography.
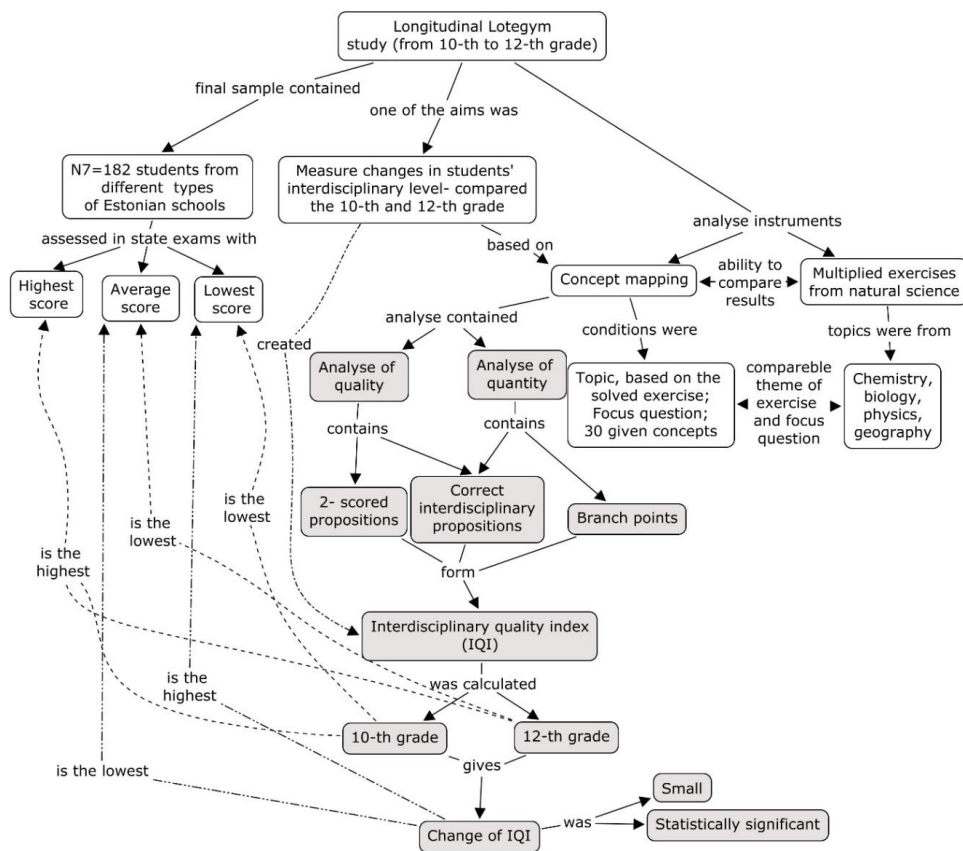
Fig. 1. Structure of the study (Note: this graph is not a concept map)

In this study, we focused on students who solved a chemistry exercise about an instant ice pack and individually constructed a corresponding concept map (N_2=343).

Results of the concept maps were compared with the results of the PISA-like test. Students were asked to create concept maps twice: initially, while in the 10th grade and secondly, while in the 12th grade. Both concept maps were constructed from the same input. The same research assistants carried out the research in the 10th and 12th grade. With the 10th grade students, assistants were asked to introduce students to concept mapping (for which they used the same presentation) and to the concept mapping software IHMC CmapTools. Research assistants gave students their individual codes to identify the maps, the focus question ("Instant ice pack"- is it only a chemistry?" the question was connected to the previously solved exercise), and 30 concepts. The concepts were selected by 85 experts (who consisted of high school teachers from different disciplines (N_{teach}=14), students from Tallinn University ($N_{univstudents}$=9) and high school students ($N_{schstudents}$=62)), and were at different abstract levels and from various subjects and topics of natural sciences and every day content: water, solubility, exothermic reaction, endothermic reaction, speed of reaction, equilibrium of chemical reaction, mole, pH, temperature of freezing, salt, energy transfer, energy, pressure, melting, friction, absorption, capillary, nerve impulse, lymphatic drainage, blood circulation, edema, dislocation, cold bag, tumor, risk, safety, pain, ethics, treatment, and first aid. The research assistants remained in the classroom during the concept map construction time (50 minutes). Students were asked not to add any new, additional concepts to the concept map. Later, tutors were asked to point out major problems that occurred (the main problem was a weak internet connection). As we wanted to compare the two concept maps constructed by the same students in 10th and 12th grade, our sample decreased to N3=182 students because some students had moved to another school, missed the session, etc. These 182 students had solved scenario-based exercises in the 10th and 12th grades, and made two concept maps (with the same focus question and pre-given concepts). Concept maps were assessed using the computer programs CmapAnalysis, MS Excel and SPSS (t-test and ANOVA).

Table 1
Division of the research's sample

Number of students	Solved scenario based multidimensional exercises (Soobard et al., 2015; Kask et al., 2015; Laius et al., 2016)	Concept map(s) constructed
N_1= 1614	**Pre-test:** two exercises on different subjects (biology, chemistry, physics or geography) Year 2011/2012	**Pre-concept map** on biology, chemistry, physics or geography (30 pre-given concepts, concepts came from 4 pre-defined clusters + focus question)
N_2= 343	**Pre-test**: two exercises on different subjects, one from chemistry and the other from biology, physics or geography Year 2011/2012	**Pre-concept map** on chemistry (30 pre-given concepts, concepts came from 4 pre-defined clusters + focus question)
N_3= 182	**Pre-test and post-test:** both of the tests contained: two exercises on different subjects, one from chemistry and the other from biology, physics or geography Years 2011/2012 and 2014	**Pre- and post-concept maps** from chemistry (30 pre-given concepts, concepts came from 4 pre-defined clusters + focus question)

Note. The sample for this study is the last row of the table

3.3. Nature of the interdisciplinary quality index of the study

The assessment of the concept maps consisted of various steps that are described in more depth in our previous studies (Soika & Reiska, 2014a; Reiska & Soika, 2015). Eighty-five experts classified the concepts into four different discipline-based clusters. The decision of whether a proposition was interdisciplinary or not was made by nature of the two concepts involved: if the concepts in the proposition were from the same cluster we defined the proposition as a *disciplinary proposition*; if the concepts were from different clusters we defined it as an *interdisciplinary proposition*.

Branch points were calculated with the CmapAnalysis program: a branch point was defined as a concept with more than two connections with another concept.

Two experts evaluated the correctness of the propositions as: (a) Propositions with a score of 2 (2-scored proposition) were deemed as high-quality and absolutely correct propositions, for example: *melting process is endothermic*; (b) Propositions with a score of 1 (1-scored proposition) are medium-quality, daily used or somehow not a correct proposition, for example: *first aid is given with cold bag*; (c) Propositions with a score of 0 (0-scored) were wrong or misunderstood propositions, for example: *melting is mainly exothermic reaction*. Whenever there were disagreements among the experts, the proposition was re-evaluated until a consensual decision was reached.

We make a distinction between a *disciplinary proposition* and an *interdisciplinary proposition*. A *correct interdisciplinary proposition* is one where concepts from different clusters are linked to together and the proposition itself has a correct meaning. Example 1: concepts *pH* and *solubility* are connected with linking phrase *depends of acid*. *pH* and *solubility* are defined by experts as concepts from chemistry. So, the proposition itself is correct, but it is a *disciplinary proposition*, it is not an *interdisciplinary proposition*. Example 2: concepts *nerve impulse* and *reaction speed* are connected by students with linking phrase *depends on*. The proposition is correct, and the experts determined that these concepts are from different clusters, because usually they are studied in different disciplines. So, this is a *correct interdisciplinary proposition* that shows that the student is able to connect concepts from different subjects. If a student makes many connections involving concepts from different disciplines (from chemistry, physics, biology, etc.) we can conclude that he or she is able to create connections between different subjects. We could say that the student possesses interdisciplinary knowledge and competences.

The interdisciplinary quality index IQI for each of the students' pre- concept maps was calculated (see Fig. 4), taking into account both quality and quantity measures of the concept map. We made the assessment of the concept map's interdisciplinarity as computer-based and easy as possible. A high-score IQI reflected well-structured, correct and interdisciplinary propositions in the concept map. We refer to these concept maps as showing a high interdisciplinary understanding, or for short, high IQI.

Determining the calculation of the IQI took several refinements. Initially, we tested other interdisciplinarity calculation methods based on a scientific literacy quality index (Soika & Reiska, 2014a). We also tried taking into account only all interdisciplinary (IQIpre) propositions, but it gave high scores to "star" shaped concept maps, and these were not well-structured concept maps and did not show interdisciplinarity. Next, we separated the IQI calculation into two parts that would assess both the quality of the concept map and the structural aspect of the concept map. We proposed, based on the nature of the concept map, that the structural measure of an interdisciplinary concept map be the ratio of the sum of interdisciplinary propositions and branch points in the map to the sum of maximum interdisciplinary propositions and

branch points for the created concept maps; and the quality measure of the concept map as the ratio of 2-scored propositions in the concept map to the maximum 2-scored propositions from created concept maps, for the given input (see equation 1).

$$IQI_1 = \underbrace{\frac{(\sum interdiscipl.\,propositions + \sum branch\,points)}{(Max_{interdiscipl.propositions} + Max_{branch\,points})}}_{\text{Quantity of the concept map}} + \underbrace{\frac{\sum 2scored\,propositions}{Max_{2scored\,propositions}}}_{\text{Quality of the concept map}}$$

(1)

Two experts found that with such calculations (equation 1), it was possible to identify highly-structured (we considered a highly-structured map to be one that is not a set of linear propositions chains or concept pairs, but instead contains 1 to 3 networks of concepts and few (1-3) or no orphan concepts) and concept maps with correct propositions. They analyzed 10 random concept maps from 5 different IQI valued groups and concluded that the calculations did not identify maps with a high interdisciplinary approach. As there is no interdisciplinary knowledge without discipline-based knowledge, the next step was to take into account both types of propositions (interdisciplinary and disciplinary propositions). We looked at the ratio of interdisciplinary and disciplinary propositions and added it to the results of branch-points and 2-scored propositions described above. Experts concluded that this method identified star-shaped concept maps. It seemed that it did not bring out interdisciplinary concept maps.

We returned to the formula presented as equation 1, but separated the sum of interdisciplinary propositions from the sum of branch points, as branch points are themselves an indicator of the structure quality of concept maps and only took into account correct interdisciplinary propositions (equation 2).

$$IQI = \underbrace{\frac{\sum correct\,interdisciplinary\,concepts}{Max\,correct\,interdisciplinary\,concepts}}_{\text{Quantity and quality of the concept map}} + \underbrace{\frac{\sum branch\,points}{Max\,branch\,points}}_{\text{Quantity of the concept map}} + \underbrace{\frac{\sum 2\,scored\,propositions}{Max_{2\,scored\,propositions}}}_{\text{Quality of the concept map}}$$

(2)

Experts examined 10 randomly selected concept maps from different IQI assessed groups, making their decision based on the structure and propositions of the concept map. They concluded that the method for calculating IQI described in equation 2 identified more concept maps that show an interdisciplinary understanding than when using equation 1. We used the IQI as expressed in equation 2 for our calculations. It's important to note that the sum of correct interdisciplinary propositions by itself is not a good measure of interdisciplinary understanding; a proper concept map structure is also important, and thus the use of the branching points count as a measure of structure.

Calculations of the IQI took into account the measures for both the individual students' concept maps and the maximum value for that measure among all maps (in 2011 and 2014). For example: there are many ways to connect a set of concepts that result in a different number of branch points. To assess the branching points in a specific concept map, we calculated how many less branch-points in the map comparted to the number in the concept map with highest number of branch points.

As the IQI consists of three different measures of quality and quantity (as shown in equation 2) the maximum IQI value is 3, as each of the 3 components can have a maximum value of 1. The maximum IQI score was considered as the highest score for both 12th and 10th grade, as we wanted we wanted to have comparable results of IQI for the 10th and 12th grade.

Fig. 2 and Fig. 3 show an example of the IQI for the pre-concept map and post-concept map for the same student. Fig. 6 shows the change in the average value of the IQI between 10th and 12th grades.
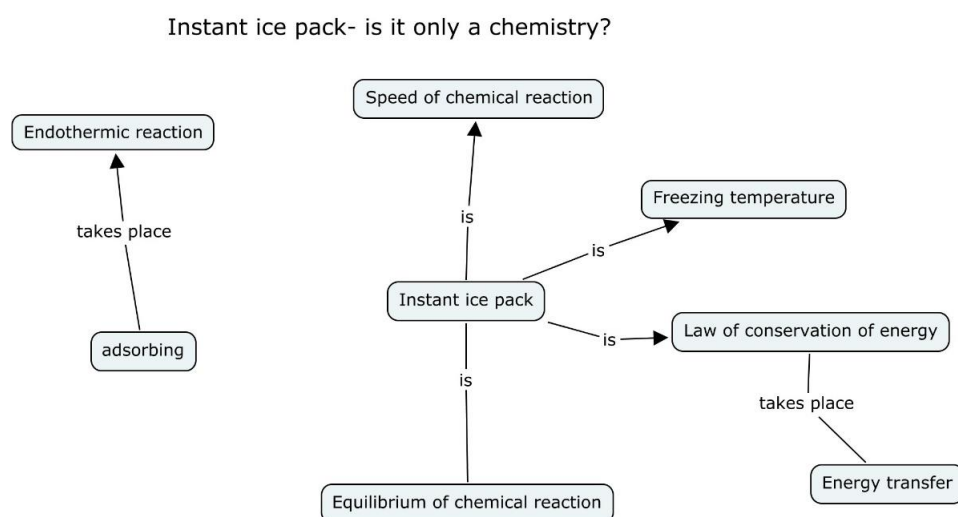

Fig. 2. Pre-concept map (made in the 10th grade) with 1 branch point (concept map is translated from Estonian), with an IQI= 0,46.
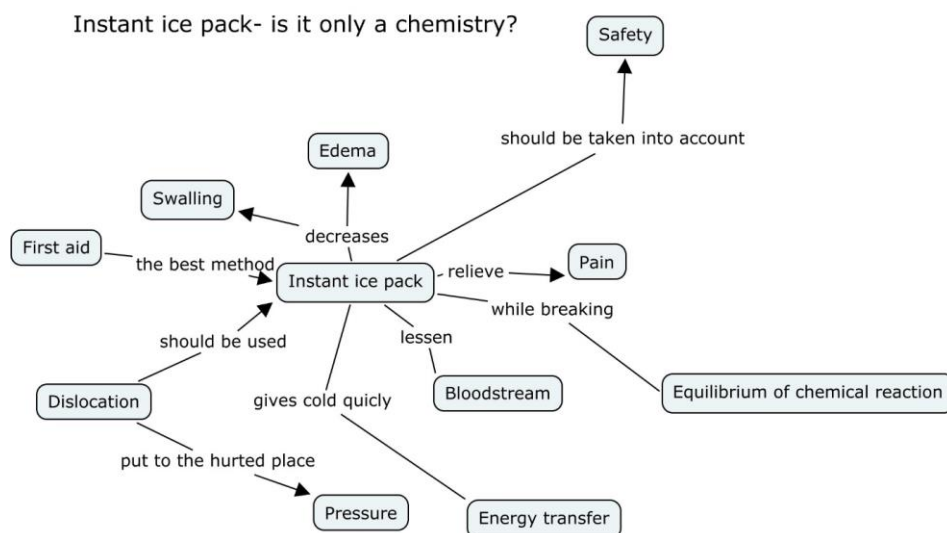
Fig. 3. Post-concept map made in 12th Grade by the same student who made the map in Fig. 2, also with 1 branch point, but with more propositions than in the students' pre-concept map (concept map is translated from Estonian), with an IQI= 0,84.

3.4. Data analysis and results

In the 10th grade, the average interdisciplinary index was 0,84 and the maximum value for IQI was 2,78. In the 12th grade, the average IQI was 0,98 and the maximum value for IQI was 2,84. The distribution of IQI values in the 10th and in the 12th grades are illustrated in Fig. 4.
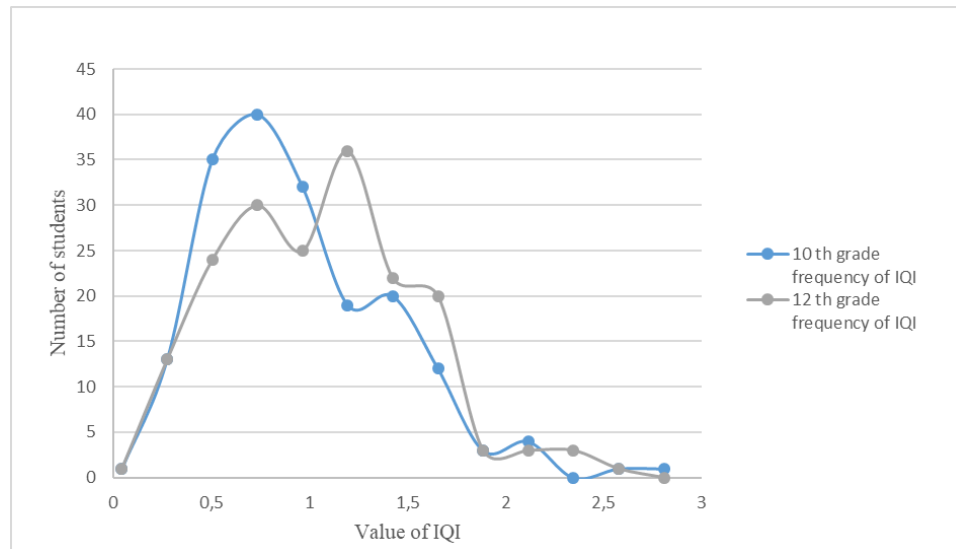

Fig. 4. IQI distribution for the 10th and 12th grades

There was some improvement in the students' concept maps from 10th to 12th grades, and though most students did not make considerably differently structured concept maps in the 12th grade, the changes were statistically significant, as presented below. Fig. 2 and Fig. 3 illustrate the improvement in one student's concept maps.

Since the sample used for the study included students from schools (a) with very good results on state exams; (b) with average results on state exams; and (c) with low results on state exams, we divided the students into those three groups:

1) students from schools with excellent results in state exams (N4 =108);
2) students from schools with average results in state exams (N5 =42);
3) students from schools with the lowest results in state exams (N6 =32).

We compared the IQI results with those from the schools' state exams. It appeared that the students from the different types of schools created differently structured concept maps. Students who studied at a school with excellent results in the state exams had the highest values for the IQI and for its 3 composite measures (the 3 components of the IQI calculation) in both 10th and 12th grade. The measures and IQI improved as they reached the 12th grade. On the other hand, the results for students who studied in schools with the lowest results on state exams had lower IQI and its composite measures, but their improvement from 10th to 12th grade was larger (these results are shown in Table 2). Additionally, students from schools with the lowest results in state exams improved the most throughout the years, but they did not reach the level that the high-scoring schools' students were at when they constructed their first concept map. Fig. 5 illustrates this statement.
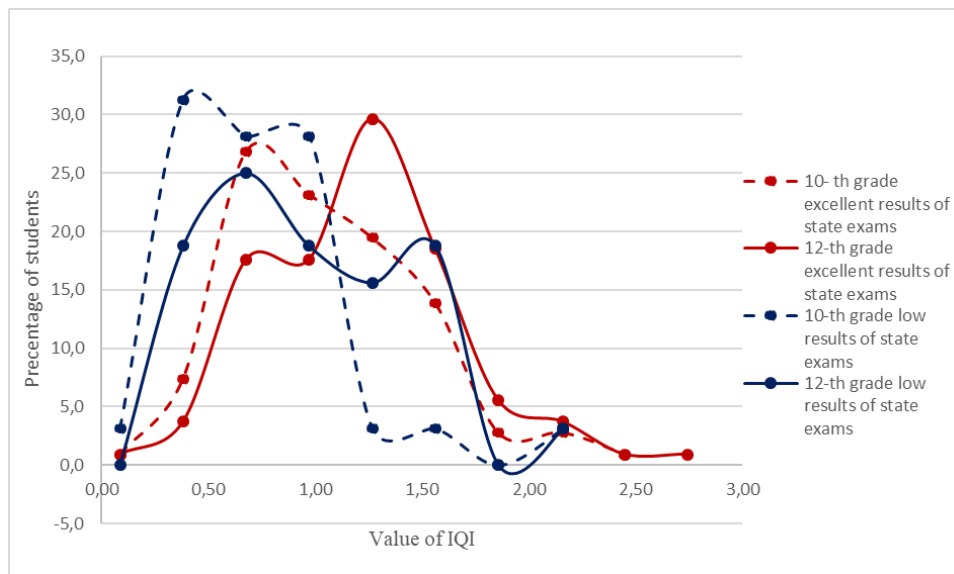
Fig. 5. Changes in IQI at schools with different results in state exams

To understand differences of the average IQI values and their distribution, we need to further examine the changes in the students' concept maps according to their school, as presented in Table 2.

Table 2
Average values of concept map measures in students' concept maps in the 10th and 12th grade (*Avrg - means average)

Results of state exams	Avrg branch points in the 10th grade	Avrg branch points in the 12th grade	Avrg 2-scored prop. in the 10th grade	Avrg 2-scored prop. in the 12th grade	Avrg Interdis. correct prop. in the 10th grade	Avrg Interdis. correct prop. in the 12th grade	Avrg interdis. quality index in the 10th grade	Avrg interdis. quality index in the 12th grade	Avrg prop. count in the 10th grade	Avrg prop. count in the 12th grade	Avrg wrong prop.in the 10th grade	Avrg wrong prop. in the 12th grade
Excellent ($N_4 = 108$)	3,87	4,28	3,84	4,74	6,69	8,88	0,94	1,07	19,18	22,53	1,84	1,69
Average ($N_5 = 42$)	3,38	2,90	2,62	3,55	5,71	6,98	0,79	0,84	17,52	18,14	2,45	1,50
Low ($N_6 = 32$)	3,00	3,31	1,63	3,38	3,56	7,06	0,59	0,86	12,34	19,19	1,63	1,69
Grand Total	3,58	3,78	3,15	4,21	5,91	8,12	0,84	0,97	17,59	20,93	1,95	1,64

Note. Prop. stands for proposition; Interdis. stands for Interdisciplinary.

Table 3
Percentage of students and their opinions of concept mapping

		Excellent results of state exams	Average results of state exams	Low results of state exams
Gender	Girls (%)	57,4	61,9	62,5
	Boys (%)	42,6	38,1	37,5
Frequency of concept map construction	more than 10 (%)	40,7	28,6	31,3
	6-10 (%)	21,3	14,3	18,8
	2-5 (%)	35,2	40,5	37,5
	only once (%)	0,0	7,1	3,1
	I have never made any (%)	2,8	7,1	6,3
Did you have any difficulties during the construction of the concept maps?	I did not have any (%)	35,2	21,4	53,1
	Generally, no (%)	55,6	57,1	34,4
	I had some problems (%)	8,3	19,0	9,4
	I had huge problems (%)	0,9	0,0	0,0
Did you enjoy constructing the concept maps?	I enjoyed it a lot (%)	3,7	4,8	12,5
	I enjoyed it (%)	47,2	26,2	40,6
	I did not enjoy it (%)	49,1	66,7	43,8
	I did not like it at all (%)	0,0	0,0	0,0
How have you constructed concept maps?	With computer and paper (%)	40,7	16,7	28,1
	Only with paper (%)	56,5	76,2	64,6
	Only with computer (%)	0,0	0,0	0,0
	I have never made any (%)	2,8	7,1	6,3
Number of schools		14	10	7
Number of students		108	42	32

Note. * These questions were asked in the 10th grade. Here are only answers from the students whose concept maps were compared in the 10th and 12th grade.

Table 2 shows that improvements tend to be larger in students from schools where the results in the state exams were the lowest. Students from schools with the highest results in state exams tended to have the highest scores for the various measures, but their improvement was not as large from 10th to 12th grade. An ANOVA test pointed out that changes in IQI comparing students from the different school groups are statistically significant in the 10th grade p= 0,009 and in the 12th grade p=0,012 (presented in Table 4).

Table 4
Statistical analysis – Significance of IQI groups

Value		*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
IQI_{10th grade}	Between Groups	2,803	2	1,401	4,839	0,009
IQI_{12th grade}	Between Groups	2,046	2	1,023	4,502	0,012

Table 3 points out how the student samples from the different types of schools differ a bit from each other. However, it seems that differences in the concept maps from students in different schools are not caused by comprehensive differences in experience or knowledge in creating concept maps or gender (Table 3). At the same time, we observe that students in schools with excellent results on state exams have made concept maps (with computer and pen) more often than students in other schools (Table 3). Table 2 shows that the students' concept maps didn't change much after two years. Students from schools with average results of state exams have not made concept maps as often as others and they did not enjoy constructing them during the study. Based on Table 3 we could have expected to obtain the most similar concept maps would be made from the excellent and low achieving schools, because their students have similar experience with concept mapping, but concept maps were created better in excellent schools and changes after the years were larger in low achieving schools. Thus, we could conclude that previous experience or knowledge in the construction of concept maps did not have an effect on the concept maps the students built.

We analyzed the changes in the three components of the IQI after the years of the study, as shown in Fig. 6. It shows that the main changes in IQI are not caused mainly by an improvement in the concept maps' structure, but by an increase in the quality of the content of the concept maps: the quality of propositions has changed more than the structure. Although the differences seem to be low, a t-test pointed out that changes in interdisciplinary correct propositions, 2-scored propositions and interdisciplinary index are statistically significant (p from 0,0002 to 0,002, presented in Table 5).

This seems to suggest that the students' understanding of concept maps and ability to construct them seems to be the same at 10th and 12th grade, since the changes in the IQI values and other concept map measures are not large (the results are shown with Fig. 6).

We looked further into what kind of changes took place in the students' interdisciplinary propositions. In the 10th grade, the averages were 3,2 high-scored propositions, 10,2 medium-scored propositions and 4,3 low-scored propositions per student. In the 12th grade the average propositions per student were: 4,0 high-scored propositions, 11,7 medium-scored propositions and 3,9 low-scored propositions.

Table 5
Statistical analysis for IQI and its' measures

| Value			t-test for Equality of Means			95% Confidence Interval of the Difference	
	t	*df*	*Sig. (2-tailed)*	*Mean Difference*	*Std. Error Difference*	*Lower*	*Upper*
Number of interdisciplinary propositions	-3,256	362	0,0012	-1,769	0,5433	-2,838	-0,701
Number of 2-scored propositions	-3,079	362	0,0022	-1,055	0,3426	-1,729	-0,381
IQI$_{final}$	-3,774	362	0,0002	-3,681	0,9753	-5,599	-1,763
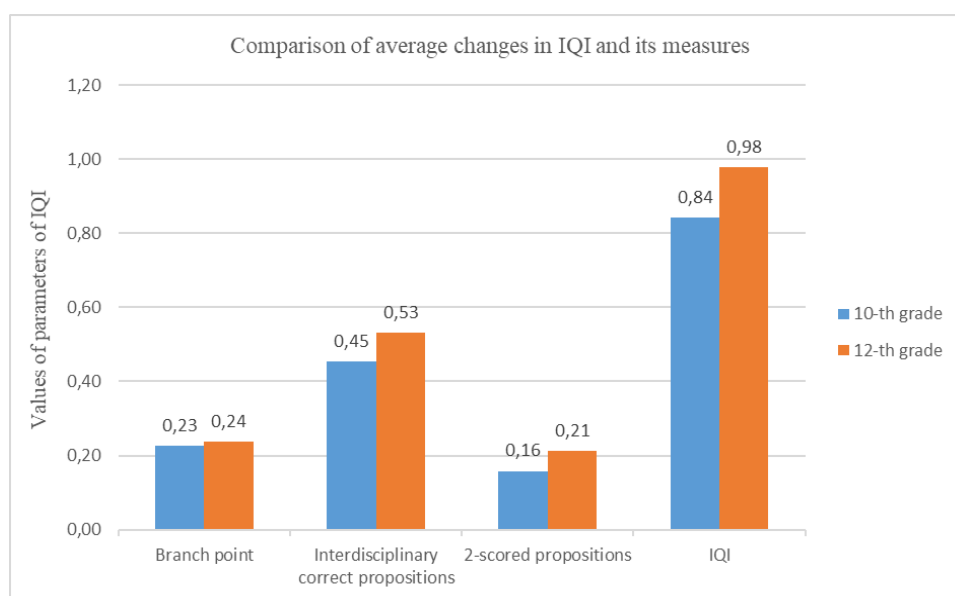
Fig. 6. Average changes in the concept maps (IQI and its measures)

In addition to the IQI calculation, where we looked at interdisciplinary propositions per students, we also analyzed same-discipline propositions as illustrated in Fig. 7. For this, we divided the propositions into four different clusters:

a) propositions involving concepts between a subject area and everyday life (for example chemistry and everyday life);

b) propositions involving concepts from different subject areas (for example proposition with concepts from chemistry and physics);

c) propositions involving concepts in the same subject areas (for example both concepts from chemistry);

d)   propositions involving everyday life concepts.

Fig. 7 shows that the largest change occurred in propositions that are created between concepts from everyday life. The smallest change in propositions appears between concepts from different subject areas (measures are statistically important).
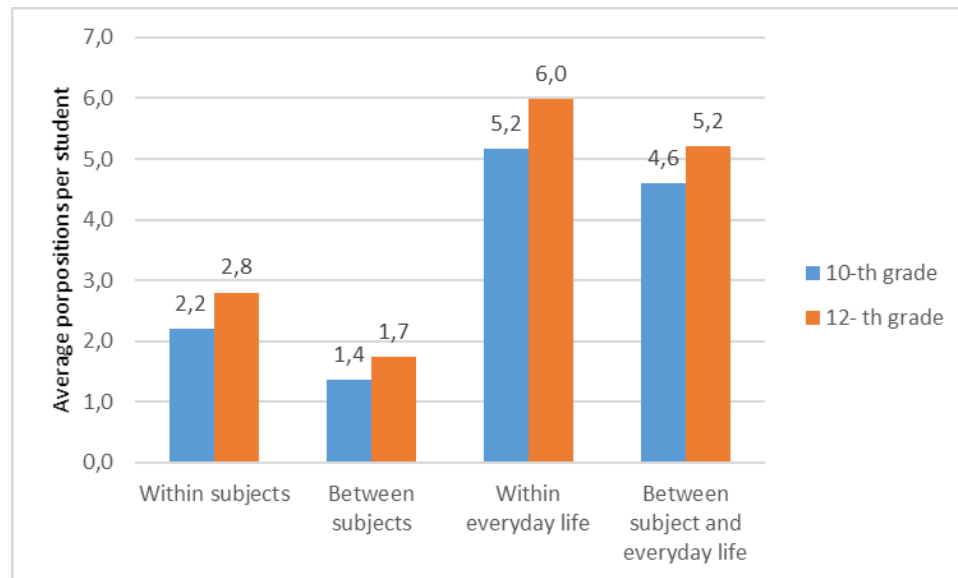
Fig. 7. Changes from 10th to 12th grade in students' interdisciplinary and disciplinary correct propositions (2-scored and 1-scored)

We also observed that the largest change appeared within propositions which scored with 1-point and connected everyday life and subject-specific concepts. There were not huge changes within 0-scored (wrong) propositions, which means that students have acquired new knowledge, but they have not removed or clarified misconceptions.

The analysis of different concept map measures (branch points, proposition count, 2-scored propositions, wrong propositions etc.) and the results of the PISA-like multidimensional test show that the largest correlated measures (r= 0,31) were the sum of correct interdisciplinary propositions of different subjects (sum of propositions included 2, 1 and 0 scored sentences) and the sum of the multidimensional test score (it means that the points from all scenario-based exercises were summed). The multidimensional test was marked by experts and it contained mainly open-ended and problem-solving exercises (Soobard et al., 2015; Kask et al., 2015, Laius et al., 2016). The result of the test is said to reflect students' knowledge and skills. The fact that the number of correct interdisciplinary propositions in the concept maps correlates with the test results seems to further validate the use of concept mapping as an assessment tool when the proper measurements of the concept map are used.

4. Discussion and conclusions

Students should be able to connect principles, concepts and theories of different subject areas and everyday life. We need instruments for evaluating interdisciplinary

comprehension, and there is a lack of such assessment instruments. Borrego with her team (2009) suggested that concept mapping could be a valuable tool for assessing interdisciplinarity. We agree that concept mapping is a valuable tool for assessing interdisciplinary understanding and analyzing levels of scientific literacy, and this study seems to demonstrate the viability of the approach. However, how to separate subject-based and interdisciplinary concepts and propositions is a concern. At the same time, we note that the concept maps constructed by the students reflect the results from state exam exercises, since schools with higher results in state exams created better concept maps than students from schools with lower results in the state exams. The results seem to suggest that more research needs to be done on the use of concept mapping for assessing interdisciplinarity.

Borrego and her team (2009) argue that using concept mapping for assessing interdisciplinarity is time consuming and causes different opinions (and discussions) between experts. We agree that the assessment of concept map is time consuming, but propose that by using a measure such as IQI, the assessment process is much faster. Borrego et al. (2009) also noticed that some students made no progress throughout their learnings, and the same phenomenon appeared in our study: in some cases, concept maps in the 12th grade were not as good as those in the 10th grade. This could occur because of the sensibility of concept mapping, which can also reflect students' feelings, instructions given, etc. (Novak, 2010; Reiska & Soika 2015; Soika & Reiska, 2014a; Anohina-Naumeca, 2015; Ruiz-Primo et al., 2001). E.g., some students who had made concept maps in the 10th grade may have not enjoyed the process, or were in a hurry in the 12th grade, but in the 10th grade had more time for creating a network of the concepts.

Borregos' team (2009) investigated concept maps with a rubric that contained three different measures (Besterfield-Sacre, Gerchak, Lyons, Shuman, & Wolfe, 2004): comprehensiveness (student's ability to define the subject area, level of knowledge of the area, and the breadth and depth of that knowledge), organization (the student's ability to systematically arrange the concepts, the hierarchy of placement and the connections/integration of the branches) and correctness (accuracy of the material presented). To evaluate the concept maps they used experts who had to assess the concept maps in a scale from 1 to 3. Sometimes the experts' opinions differed from each other. We developed the IQI such that there is no need for experts to agree except for the assessment of the quality of the propositions. IQI consists of three components that bring out the quality (correct interdisciplinary propositions + 2-scored proposition count) and quantity (structure) components of the concept map (correct interdisciplinary propositions + count of branch points) as is illustrated in equation (2). Given the size of our sample, we decreased the subjective aspect of the assessment and made the process as computer-based as possible. It would have been very time-consuming to have experts assess 364 concept maps manually. The IQI developed provides a measure of interdisciplinarity in the map's propositions and structure without the need to manually assess each concept map.

Our first aim was to investigate how students' interdisciplinarity understanding changes throughout the high school studies. Using the IQI we developed, we analyzed how the IQI and its components changed through the years of the study. The study showed that students were able to create more high-scored correct propositions using concepts from the same cluster than they made with concepts from different clusters, which seems to suggest that in Estonian schools studying within subjects is stronger than across subject areas. The study also showed that students were better able to construct propositions between familiar daily-used concepts than propositions between concepts from different fields.

A second aim was to compare the differences and changes in interdisciplinarity among students from different schools. In Estonia, students have to pass entrance tests to some high schools. As a result, the achievement in these tests is available for the different schools. We didn't expect that the structure and changes in the concept map varied by the ranking (based on the entrance tests) of schools. Students from schools with higher results at state exams often had complex concept maps in the 10th grade, but their post-concept maps in the 12th grade had not changed as much with respect to their 10th grade maps as concept maps of those students who studied at schools where students' achievements are not as high. Additionally, students from schools with lower state exams level did not reach in 12th grade the level of concept maps measures of the maps of 10th grade students from schools with better scores at national examinations. This result could be explained by the nature of the learning curve. We noticed that concept maps of students who are studying at a school where average state exams results are high did not improve much throughout their studies at high school. Probably some of the students are closer to the "expert" level (based on competences that are defined at the state curriculum or the expectations in the classroom) and therefore they did not develop as much as students at schools where the average result of national exams is lower. We could suppose that most of the students at these schools are not as close to experts of some subject, and are therefore in the phase of active learning. Their concept maps would then show more improvement in the 12th-grade maps as they increase their understandings of the subject areas.

On a more general level, educators and scientists need to find ways to better support the talent and achievement of high school students to overall improve their development. We need to find out how to improve the curriculum so that students reach a deeper understanding of the various disciplines and a higher degree of interdisciplinarity when they graduate from school.

Throughout the study we evaluate the feasibility of automatically assessing a large number of concept maps to measure level of interdisciplinarity understanding expressed by the map builders. We believe that this study increased our understanding of the reliability of concept mapping as an assessment tool, because we had the opportunity to compare the results of a reliable PISA-like test and the results of a qualitative and quantitative assessment of concept maps. Finally, we realize that assessment of interdisciplinarity with concept mapping is promising, but more research is needed before we can reach further conclusions.

5. Limitations

The changes in the number of students throughout the study made it difficult to evaluate the results, but this is a problem with longitudinal studies where students are moving, and the sample is affected.

References

Adler, P. S., & Clark, K. B. (1991). Behind the learning curve: A sketch of the learning process. *Management Science, 37*(3), 267–281.

Angelo, T. A. (1995). Reassessing (and defining) assessment. *AAHE Bulletin, 48*(3), 7–9.

Anohina-Naumeca, A., Grundspenkis, J., & Strautmane, M. (2011). The concept map-based assessment system: Functional capabilities, evolution, and experimental results. *International Journal of Continuing Engineering Education and Life Long Learning,*

21(4), 308–327.

Anohina-Naumeca, A. (2015). Justifying the usage of concept mapping as a tool for the formative assessment of the structural knowledge of engineering students. *Knowledge Management & E-Learning, 7*(1), 56–72.

Arrow, K. J. (1962). The economic implications of learning by doing. *The review of economic studies, 29*(3), 155–173.

Austin, L. B., & Shore, B. M. (1995). Using concept mapping for assessment in physics. *Physics Education, 30*(1), 41–45.

Ausubel, D. P. (1968). *The psychology of meaningful learning: An introduction to school learning*. New York, NY: Grune and Stratton.

Benzel, E. C., & Orr, R. D. (2011). A steep learning curve is a good thing! *The Spine Journal, 11*(2), 131–132.

Besterfield-Sacre, M., Gerchak, J., Lyons, M., Shuman, L. J., & Wolfe, H. (2004). Scoring concept maps: An integrated rubric for assessing engineering education. *Journal of Engineering Education, 93*(2), 105–115.

Biological Sciences Curriculum Study (BSCS). (1993). *Developing biological literacy: A guide to developing secondary and post-secondary biology curricula*. Dubuque, IA: Kendall Hunt Pub Co.

Borrego, M., Newswander, C. B., McNair, L. D., McGinnis, S., & Paretti, M. C. (2009). Using concept maps to assess interdisciplinary integration of green engineering knowledge. *Advances in Engineering Education, 1*(3), 1–26.

Bretz, S. L. (2001). Novak's theory of education: Human constructivism and meaningful learning. *Journal of Chemical Education 78*(8), 1107–1117.

Bybee, R. W. (1997). Toward an understanding of scientific literacy. In W. Gräber & C. Bolte (Eds.), *Scientific Literacy: An International Symposium* (pp. 37–68). Kiel, Germany: Institute for Science Education (IPN).

Cañas, A. J., Bunch, L., & Reiska, P. (2010). CmapAnalysis: An extensible concept map analysis tool. In *Proceedings. of the Fourth International. Conference on Concept Mapping* (pp. 75–82). Viña del Mar, Chile: Universidad de Chile.

Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., ... Carvajal, R. (2004). CmapTools: A knowledge modeling and sharing environment. In *Proceedings of the first International Conference on Concept Mapping* (Vol. 1, pp. 125–133). Pamplona, Spain.

Cañas, A. J., Novak, J. D., & Reiska, P. (2015). How good is my concept map? Am I a good Cmapper? *Knowledge Management & E-Learning, 7*(1), 6–19.

Cañas, A. J., Reiska, P., & Möllits, A. (2017). Developing higher-order thinking skills with concept mapping: A case of pedagogic frailty. *Knowledge Management & E-Learning, 9*(3), 348–365.

Emenike, M. E., Danielson, N. D., & Bretz, S. L. (2011). Meaningful learning in a first-year chemistry laboratory course: Differences across classical, discovery, and instrumental experiments. *Journal of College Science Teaching, 41*(2), 89–97.

Haridus- JA Teadusministeerium. (2018). *Eesti elukestva õppe strateegia 2020*. Retrieved from https://www.hm.ee/et/elukestva-oppe-strateegia-2020

Henno, I. (2015). *Loodusteaduste õppimisest ja õpetamisest Eesti koolides rahvusvaheliste võrdlusuuringute taustal*. Doctoral dissertation, Tallinn University, Estonia.

Holbrook, J., & Rannikmäe, M. (2009). The meaning of scientific literacy. *International Journal of Environmental and Science Education, 4*(3), 275–288.

Ivanitskaya, L., Clark, D., Montgomery, G., & Primeau, R. (2002). Interdisciplinary learning: Process and outcomes. *Innovative Higher Education, 27*(2), 95–111.

Kask, K., Ploomipuu, I., & Rannikmäe, M. (2015). Changes in cognitive skills during a gymnasium chemistry course. *Procedia-Social and Behavioral Sciences, 177*, 367–

371.

Kinchin, I. M. (2011). Visualising knowledge structures in biology: Discipline, curriculum and student understanding. *Journal of Biological Education, 45*(4), 183–189.

Kinchin, I. M. (2015). Editorial: Novakian concept mapping in university and professional education. *Knowledge Management & E-Learning, 7*(1), 1–5.

Kinchin, I. M. (2017). Pedagogic frailty: A concept analysis. *Knowledge Management & E-Learning, 9*(3), 295–310.

Klassen, S. (2006). Contextual assessment in science education: Background, issues, and policy. *Science Education, 90*(5), 820–851.

Laius, A., Post, A., & Rannikmäe, M. (2016). Assessment of scientific literacy of Estonian gymnasium students during the operation of competence-based science curriculum. *Universal Journal of Educational Research, 4*(5), 1142−1147.

Mansilla, V. B., & Duraisingh, E. D. (2007). Targeted assessment of students' interdisciplinary work: An empirically grounded framework proposed. *The Journal of Higher Education, 78*(2), 215–237.

Miller, N. L. (2008). *An exploration of computer-mediated skill acquisition in concept mapping by in-service Panamanian public elementary schoolteachers*. Doctoral dissertation, Universitat Oberta de Catalunya, Spain.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

Ngwenyama, O., Guergachi, A., & McLaren, T. (2007). Using the learning curve to maximize IT productivity: A decision analysis model for timing software upgrades. *International Journal of Production Economics, 105*(2), 524–535.

Next Generation Science Standards (NGSS). (2013). *Science education in the 21st century: Why K–12 science standards matter—and why the time is right to develop next generation science standards*. Retrieved from http://www.nextgenscience.org/sites/ngss/files/NGSS%20fact%20sheet.pdf

Nissani, M. (1997) Ten cheers for interdisciplinarity: The case for interdisciplinary knowledge and research. *The Social Science Journal, 34*(2), 201–216.

Novak, J. D. (2010). *Learning, creating and using knowledge: Concept maps as facilitative tools in schools and corporations*. New York, NY: Routledge.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge, UK: Cambridge University Press.

OECD. (2016). *PISA 2015 results (Volume I): Excellence and equity in education*. OECD Publishing, Paris. doi: http://dx.doi.org/10.1787/9789264266490-en

Passerotti, C. C., Franco, F., Bissoli, J. C., Tiseo, B., Oliveira, C. M., Buchalla, C. A., ... Nguyen, H. T. (2015). Comparison of the learning curves and frustration level in performing laparoscopic and robotic training skills by experts and novices. *International Urology and Nephrology, 47*(7), 1075–1084.

Reiska, P., & Soika, K. (2015). Suggestions for teacher education from concept mapping studies. *Knowledge Management & E-Learning. 7*(1), 149–161.

Ritter, F. E., & Schooler, L. J. (2001). The learning curve. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 8602–8605). Amsterdam: Pergamon Press.

Ruiz-Primo, M. A. (2004). Examining concept maps as an assessment tool. In *Proceedings of the first International Conference on Concept Mapping*. Pamplona, Spain.

Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (2001). Comparison of the reliability and validity of scores from two concept-mapping techniques. *Journal of*

Research in Science Teaching, 38(2), 260–278.

Ruiz-Primo, M. A., Schultz, S. E., & Shavelson, R. J. (1997). *Concept map-based assessment in science: Two exploratory studies*. Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching, 33*(6), 569–600.

Schaal, S., Bogner, F. X., & Girwidz, R. (2010). Concept mapping assessment of media assisted learning in interdisciplinary science education. *Research in Science Education, 40*(3), 339–352.

Schwendimann, B. A. (2014). Multi-level analysis strategy to make sense of concept maps. In *Proceedings of Sixth International Conference on Concept Mapping* (pp. 363–369). Santos, Brazil.

Soika, K., & Reiska, P. (2013). Large scale studies with concept mapping. *Journal for Educators, Teachers and Trainers. 4*(1), 142–153.

Soika, K., & Reiska, P. (2014a). Using concept mapping for assessment in science education. *Journal of Baltic Science Education. 13*(5), 662–673.

Soika, K., & Reiska, P. (2014b). Assessing students' cognitive skills with concept mapping. In *Proceedings of 8th International Technology, Education and Development Conference*.

Soika, K., & Reiska, P. (2014c). Assessing students' interdisciplinary approach with concept mapping. In *Proceedings of Sixth International Conference on Concept Mapping* (pp. 71–79). Santos, Brazil.

Soobard, R., Rannikmäe, M., & Reiska, P. (2015). Upper secondary schools students' progression in operational scientific skills – A comparison between grade 10 and 12. *Procedia - Social and Behavioral Sciences, 177*, 295−299.

Stowe, D. E., & Eder, D. (2002). Interdisciplinary program assessment. *Issues in Integrative Studies, 20*, 77–101.

Tao, C. (2015). *Development of a knowledge assessment system based on concept maps and differential weighting approaches*. Doctoral dissertation, Virginia Polytechnic Institute and State University, USA.

You, H. S., Marshall, J. A., & Delgado, C. (2018). Assessing students' disciplinary and interdisciplinary understanding of global carbon cycling. *Journal of Research in Science Teaching. 55*(3), 377–398.