LUMAT

General Issue
Vol 5 No 1 (2017)

DEVELOPMENT OF AN INVENTORY FOR ALTERNATIVE CONCEPTION AMONG STUDENTS IN CHEMISTRY

Per-Odd Eggen, Jonas Persson, Elisabeth Egholm Jacobsen, Bjørn Hafskjold

Received: 2.1.2017
Accepted: 30.6.2017

Abstract A chemistry concept inventory (Chemical Concept Inventory 3.0/CCI 3.0) has been developed for assessing students learning and identifying the alternative conceptions that students may have in general chemistry. The conceptions in question are assumed to be mainly learned in school and to a less degree in student's daily life. The inventory presented here aims at functioning as a tool for adjusting teaching practices in chemistry and is mainly designed for assessing the learning outcome during university general chemistry courses. Used as a pre-test the inventory may also give information about student's starting point when entering university's first year chemistry courses.The inventory also aims at functioning as a tool for adjusting teaching practices in chemistry. It has been administered as a pre- and post-test in general chemistry courses at the Norwegian University of Science and Technology (NTNU), and evaluated using different statistical tests, focusing both on item analysis and the on the entire test. The results indicate that the concept inventory is a reliable and discriminating tool in the present context.

1. Introduction

During the last decades, studies aimed at describing the concepts held by students in the fields of Science, Technology, Engineering and Mathematics (STEM) have been performed. Concepts are developed from early age when children form intuitive ideas of natural phenomena. During the process of learning more about the natural world, by observations or a theoretical approach, they develop new or revised concepts based on their own interpretation of new information within their own context with existing ideas and beliefs. The concepts that are not consistent with the established consensus are sometimes called misconceptions (Smith et al 1994) or alternative conceptions.

Students in introductory courses in chemistry exhibit a number of alternative concepts concerning chemical behaviour. There has been published a number of reviews of common alternative conceptions in chemistry (Bowen & Bunce, 1997), (Stavy, Learning Science in the Schools, 1995) (Gabel & Bunce, 1994) (Wandersee, Mintzes, & Novak, 1994), as well as an extensive bibliography (Pfundt & Duit, 2000), dealing with these issues. This research has give valuable background knowledge for later development of assessment methods in chemistry education (Treagust & Chiu, 2011) (Cloonan & Hutchinson, 2010).

Some alternative conceptions influence the learning process in a deeper sense than just producing inadequate explanations to questions. (Nakleh, 1992), (Krajcik, 1991) The students reconstruct their conceptual understanding, in the face of new information, from their present conceptions. When encountering new information that contradicts their alternative conceptions, they might find it difficult to accept this information as it seems to be wrong. This may lead to a conflict that can be solved in different ways, the new information can be; ignored, rejected, disbelieved, deemed irrelevant, held for consideration later, reinterpreted to fit the students' current conceptions or accepted with only minor changes in the student's conception. The information may also be accepted and the prior conception revised. Chemical concepts such as a solution may be learned in daily life, but many (maybe most) such concepts must be assumed to be learned in school. Tools for mapping and analysing students' conceptions may therefore be useful when adjusting teaching towards a practice that can facilitate deeper understanding.

A number of teaching techniques have emerged during the last decades, with the aim of increasing the learning outcome among students. The new methods have demonstrated that compared to a teacher-centred lecturing, a more student-centred model of education using more hands-on and inquiry-based approach, increases the student's knowledge and conceptual understanding of a subject (Taber, 2009). In order to compare different methods it is important to have assessment tools that are able to measure the students' conceptual understanding as well as to understand what conceptions and background limitations the students have when entering a class.

The Force Concept Inventory (FCI) is an assessment tool created by Hestenes et al. (Hestenes, Wells, & Swackhamer, 1992) for use in high school and college physics classes. A concept inventory consists of a series of multiple-choice questions, based on qualitative conceptual orientated problems. It aims to measure deep understanding and conceptual knowledge rather than a student's ability to solve problems. The results can be compared to the results of students in classes with different teaching methods in order to determine teaching efficiency. In addition, a concept inventory can sample the alternative conceptions in a student population, as well as being able to assess the progression during a course by administrating pre- and post-test using the inventory.

Our aim was to develop a chemical concept inventory for use in general chemistry courses at Norwegian universities as well as in upper secondary school in Norway. The inventory serves two purposes: The primary purpose is to map students´ understanding of concepts in chemistry. The secondary is to use the inventory as an independent tool for evaluation of learning activities. When comparing a chemical concept inventory with e.g. the Force Concept Inventory, one has to bear in mind that conceptions relating to forces to a large extent is learned in daily life, but most chemical conceptions can be assumed to result from teaching practices. In this article we describe the development of the inventory and some results from its application to students at university level. The applications assured the quality of the inventory, whereas evaluating students' learning activities will have to be based on more extensive and systematic tests. This will be the topic of future papers. Together with reviews of common alternative conceptions in chemistry, concept inventories may constitute a basis for understanding students learning difficulties and achievements. The inventories may also display differences in learning outcomes when comparing students from different learning institutions.

2. Development and validity

The test was developed from a number of existing validated concept inventories in chemistry (Krause, Birk, Bauer, Jenkins, & Pavelich, 2004) (Mulford & Robinson, 2002) in addition to questions written by the authors based on literature and personal experience. The main objective is to use the inventory in the compulsory general chemistry courses for mapping students' understanding of concepts and to use it for finding effects of teaching activities. The inventory is mainly intended for use in first year of university studies, but may have a wider application, for example in upper secondary school. The CCI is, however, only evaluated and validated in a university context. From a test bank of about 113 concept questions originating from the chemistry inventories of Krause et al. (Krause, Birk, Bauer, Jenkins, & Pavelich, 2004), Mulford and Robinson (Mulford & Robinson, 2002) and about 30 questions made by ourselves, we excluded a number of questions as being out of context or not suitable for our purposes, leaving us with 70 remaining questions. These questions were compared with the curriculum of all the general chemistry courses given at NTNU, by us and other experts at NTNU, as to exclude questions not covered.

The number of questions was considered too large to administer as a single test, and a sub-set of 50 questions was used in a first version (1.0) of the inventory administered as a post-test in a course during the 2014 spring semester. The results from the test were analysed with respect to difficulty and discrimination. Nine questions were judged as unsatisfactory, either being too easy or not sufficiently discriminating. The test was also considered to be too time consuming and a number of questions of duplicate nature were removed. The next step was to construct two tests (2.0 and 2.1) with major overlap of questions while testing new questions. These tests consisting of 38 and 40 questions, respectively, were administered as post-tests in two different courses in the 2014 fall semester. The result judged three and four questions, respectively, to be unsatisfactory. Combining these results made it possible to construct an inventory (3.0) of 40 questions with an estimated completion time of 40-45 min. During the 2015 spring semester this inventory was administered to 60 students in general chemistry courses. 25 of these belonged to phys./math. master program, 18 to material technology, 13 to nano technology and four students to other programs. The chemistry course was compulsory only in the nano and phys./math. programs.

Soliciting expert opinions is a standard method of assessing the validity of a test. The term “validity”, which is not a statistical construct, refers to the extent which the test actually measures what it is supposed to measure. Validity, as such, can have different aspects (Kline, 1986). Face validity can be determined by a common sense reading of an instrument; a test would lack face validity if it tested concepts unrelated to the subject. Content validity reflects the coverage of the subject, i.e. does the test cover enough of a topic? Both of these are typically assessed by expert consensus, as has been done with this inventory. The experts assessing the test have all lectured general chemistry courses for several years at NTNU and partially also at upper secondary school, thus having insight in the relevant chemistry curricula.

The inventory described in this paper covers a wide range of topics generally introduced in an undergraduate general chemistry course, but cannot claim to cover the majority of chemical concepts. It should rather be seen as a starting point, at least in a Nordic context. The questions and the topics that are covered are given in Table 1.

Table 1. Topics covered in the Chemical Concept Inventory

Topic	**Question #**
Molecular geometry	28, 29
Atomic structure	4
Stoichiometry	24
Chemical bonding	2, 3, 26, 27, 32, 33
Gases	20, 21, 31
Chemical equilibrium	9, 10, 11, 12, 13, 14, 35
Redox reactions and electrochemistry	18, 19, 36, 38
Phase equilibrium	22, 25, 30

Thermochemistry	32, 37
Thermodynamics	1, 34, 39, 40
Intermolecular forces	5,6,7,8,
Acid-base equilibria	15, 16, 17

The sequence of these questions was in part random. We could not detect any effect of the sequence, but this may be investigated in later studies.

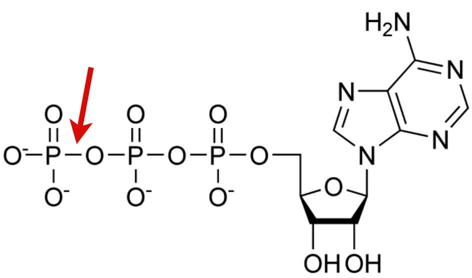
Two examples of questions from this CCI are given in Figures 1 and 2.

Which of the following statements describes a situation where the intermolecular forces are overcome?

A) Decomposition of ammonia (NH_3) into nitrogen and hydrogen
B) Decomposition of sodium chloride (NaCl) into chlorine and sodium ions
C) Evaporation of methanol molecules from liquid methanol (CH_3OH).
D) Separating hydrogen atoms from each other in hydrogen gas (H_2).
E) All of the above.

Figure 1: Question 8 from the concept inventory examining the nature of intermolecular forces.

In the picture of an ATP-molecule, a specific bond is indicated with an arrow. Which statement on this bond is CORRECT?



A) It always takes energy to break a bond. This is also true for this bond.
B) It takes energy to break most bonds, but this is a high energy bond that will release energy when broken.
C) All bonds release energy when broken.
D) It takes energy to break covalent bonds, but this is an ionic bond, which release energy when broken.

Figure 2: Question 33 from the concept inventory examining the nature of intramolecular forces.

3. Results

3.1 Student group

At the Norwegian University of Science and Technology, all engineering masters are required to study at least one chemistry course. This gives us an opportunity, not only to study the results of chemistry masters, but also to study students with different interests in chemistry. The final inventory (3.0) was given in a general chemistry course (TMT4110) with students from three different master programs. In this course, the students are supposedly high-achieving, mastering in Physics, Nanotechnology and Materials Science and Engineering. The Physics and Nanotechnology students are generally regarded as high-achieving students as admission grades are higher compared with the other master programs at NTNU. The results may therefore be expected to be higher for this specific course compared to other courses, possibly with the exception of e.g. Chemistry master students. A study of results for individual questions in the different versions of the inventory through the development phase is consistent with this hypothesis.The test, given as a post-test, was voluntary with no extra credit given. 60 students performed the test during a 45 min exercise (problem solving)-session.

3.2 Analysis of results

In order to investigate the reliability and discrimination power of the inventory, a number of statistical tests focusing both on individual items and on the test as a whole, has been performed.

There exist two aspects with test reliability; consistency and discriminatory power. A test is reliable if it is consistent within itself and over time. If a test is shown to be reliable, one can assume that the same students would get the same score if they would take the test again. A large variance in the test score of a reliable test will then depend on a systematic variation in the student population, where different levels of understanding or mastery will give different scores on the test. Both these aspects of test reliability can be assessed statistically.

Using the results from the inventory we performed five statistical tests: three focusing on individual items (item difficulty index, item discrimination index, item point biserial coefficient) and two on the test as a whole (Kuder-Richardson test reliability and Ferguson's δ). (Kline, 1986) (Kuder & Richardson, 1937)

3.2.1 Item difficulty index

The item difficulty index (P) is a measure of the difficulty of each test item. It is defined as the ratio of the total number N_1 of correct answers to the total number N of students who answered the specific item:

$$P = \frac{N_1}{N} \tag{1}$$

The difficulty index is somewhat misnamed, since it is simply the proportion of correct answers to a particular item, where the name "easiness index" might be more appropriate. The greater the P value, the higher percentage of correct answers and consequently the easier the item is for the population. The difficulty index will also depend on the population, which is the case in this study.

There are a number of different criteria for acceptable values of the difficulty index for a test (Doran 1980). The optimum value for an item would be P = 0.5, while it is useful to have a sensible range. A widely adopted criterion requires the difficulty index to be between 0.3 and 0.9 for each question. For a test with a large number (M) of items, the test difficulty may instead be considered as the average difficulty index ($\overline{P}$) of all the items (P_i):

$$\overline{P} = \frac{1}{M}\sum_{i=1}^{M} P_i \tag{2}$$

The results from the test gave a range of [0.33, 0.92] with a mean of 0.66 and only one question over the acceptable limit. This we regard as satisfactory even though the mean is somewhat high, indicating that the questions are easy, however we do not expect any ceiling effects. One will also expect that difficulties will change between different student groups as found in the case of FCI (Persson, 2015).

3.2.2 Item discrimination index

The item discrimination index (D) is a measure of the discriminatory power for individual items in a test. That is, the extent to which an individual test item distinguishes a student who know the material well from those who do not. A high discrimination index will therefore indicate a higher probability for students with a robust knowledge to answer the item correctly, while those with less knowledge or misconceptions more probably will get the wrong answer.

The item discrimination index (D) is calculated by first dividing the sample into two groups of equal size, a high (H) score group and a low (L) score group based on their individual total scores on the test. For a specific item, one counts the number of correct answers in both the high and low groups: N_H and N_L. If the total number of students taking the test is N, the discrimination index for a specific item can be calculated as

$$D = \frac{N_H - N_L}{N / K} \tag{3}$$

where K is a numerical factor based on how the division in the high and low group is made. If we split the sample in two, using the median, the high and low groups consist each of 50% of the total sample, giving K=2. However, it is possible to use other groupings, for example taking the top 25% as the high group and the bottom 25% as the low group. The 50% - 50% grouping may underestimate the discrimination power, since it takes all students into account. To reduce the probability of underestimating the discrimination power we use a 25% - 25% grouping, which necessary means that we have to discard half of the available data. The discrimination index is then expressed as:

$$D = \frac{N_H - N_L}{N / 4} \tag{4}$$

The range of the item discrimination index D is [−1,+1], where +1 is the best value and −1 the worst. If all students in the high score group and none in the low score group get the correct answer the discrimination index would be +1. If none in the high score group and all in the low score group get the correct answer the discrimination index would be −1. These extremes are unlikely, but it is a good practice to remove items with negative discrimination index. A question is typically considered to provide a good discrimination if D ≥ 0.3 (Doran 1980). In a test with a number of items it is possible to allow a few items with a lower discrimination index, but the majority should have high discrimination indices in order to ensure that the test can distinguish strong and weak mastery. It is possible to check this by calculating the averaged discrimination index ($\overline{D}$) for all items in the test with equation 5:

$$\overline{D} = \frac{1}{M}\sum_{i=1}^{M} D_i \tag{5}$$

We found that the average discrimination index where 0.45 with a range of [0.13, 0.73] for the individual questions. However, 10 questions in this students group were below the recommended value, though most of them slightly. These were questions 1: (D=0,13); 11and 31(D=0,2); 13, 19, 24, 26, 30, 33 and 37 (D=0,27). When comparing the results of these questions in our development tests we found that most of them were over the recommended limit. This indicates that that the questions are slightly less discriminating in this specific student group, which is believed to be high-achieving. As the overall result is satisfactory, the test as a whole is accepted as discriminating enough, though special attention must be paid in the case of high-achieving students.

3.2.3 Point biserial coefficient

The point biserial coefficient is a measure of the individual item reliability. It reflects the correlation between the total score and the score on individual items. A positive coefficient indicates that a student with a high total score is more likely to answer the item correctly than a student with a low total score. To calculate the point biserial coefficient for an item, one calculates the correlation between the score for a question and the total scores. The student's score on an item can have two values: 1 (correct) or 0 (wrong). If the number of items in the test is sufficiently large, >20, the test can be viewed as continuous. The point biserial coefficient can then be defined as:

$$r_{pbc} = \frac{\overline{X_1} - \overline{X_0}}{\sigma_x}\sqrt{\frac{P}{1-P}} \tag{6}$$

Where $\overline{X_1}$ is the average total score for those who answered a item correctly, $\overline{X_0}$ is the average total score for all, σ_x is the standard deviation of the total scores and P is the difficulty index for this item. A reliable item should be consistent with the whole test so a high correlation between individual questions score and the total score is desirable. A satisfactory point biserial coefficient is $r_{pbc}>0.2$. Items with lower values may still be used, as long as the number is small, but the test as a whole should have an average higher than 0.2. The average point biserial coefficients for the inventory were found to be 0.41, with only one question below the recommended limit. The range in this case was [0.13, 0.57]. The results show that the test is reliable for our purposes.

3.3 Test analysis

The reliability of single items in the test is measured by the point biserial coefficient. In order to examine the reliability of the whole test, other methods have to be used. In this work we use two measures of the reliability for the test as a whole: Kuder-Richardson reliability index and Ferguson's delta (δ) (Kuder & Richardson, 1937) (Kline, 1986).

3.3.1 Kuder-Richardson reliability index

One way to evaluate the reliability of a test is to administer it twice to the same sample. In such a case we would expect a significant correlation between the two test scores, provided the students' performance is stable and the test conditions are the same. The correlation coefficient between the two sets of scores will be defining the reliability index of the test. It is obvious that this method is not practical to use, as persons in the sample will remember the questions.

In the case of a test that has been specifically designed for a certain knowledge domain with parallel questions, the Spearman-Brown formula (Ghiselli, Campbell & Zedeck 1981) can be used as an alternative to calculate the reliability index. This equation connects the reliability index with the correlation between any two parallel equally sized subsets of the test. Kuder and Richardson took this idea further, by dividing the test into the smallest possible subsets, individual items (Kuder & Richardson, 1937). That is, each item is considered as a single parallel test assuming that the means, variance and standard deviation is the same for all items. The result gives the reliability index as:

$$r_{test} = \frac{M}{M-1}\left(1 - \frac{\sum \sigma_{xi}^2}{\sigma_x^2}\right) \tag{7}$$

where M is the number of items in the test, σ_{xi} is the standard deviation for the i^{th} item score and σ_x is the standard deviation of the total test score.

This expression takes the different variances of the items into account, relaxing the assumption that all items must have the same means, variance and standard deviation. For multiple-choice tests the formula can be rewritten as:

$$r_{test} = \frac{M}{M-1}\left(1 - \frac{\sum P_i\left(1-P_i\right)}{\sigma_x^2}\right) \tag{8}$$

where M is the number of items in the test, P_i is the difficulty index for each item and σ_x is the standard deviation of the total test score. This is the Kuder-Richardson reliability formulas, often referred to as KR-20 and KR-21 as being formula 20 and 21 in Kuder and Richardson's original paper (Kuder & Richardson, 1937). The possible range of the Kuder-Richardson reliability index is between 0 and 1, where a value greater than 0.7 would make the test reliable for group measurements and a value over 0.8 for assessing individuals (Doran

1980). In this study the obtained Kuder-Richardson reliability index of 0.88. This value indicates that our inventory is also suitable for individual assessment.

3.3.2 Ferguson's delta

Ferguson's delta is another whole test statistic. It measures the discriminatory power of the whole test by investigating how the students' scores are distributed. One aims at a broad distribution in total scores, as this indicates a better discrimination.

The expression of Ferguson's delta can be written as:

$$\delta = \frac{N^2 - \sum f_i^2}{N^2 - \left(N^2/(M+1)\right)} \tag{9}$$

where N is the number of students taking the test, M is the number of items in the test and f_i is the frequency of cases with the same score. If a test has a Ferguson's delta greater than 0.90, it is considered to provide a good discrimination among students (Kline 1986, p. 144 and 150). In our study the Ferguson's delta was 0.98.

4. Discussion

There is little knowledge on the effectiveness of different teaching methods in chemistry, and as demonstrated by Mazur (Mazur, 2009) and others, university grades do not always reflect subject knowledge satisfactory. According to John Hattie, most reports from tests of teaching methods are positive, but only a fraction of these methods are in the "zone of desired effects" (Hattie, 2009). Hence, here seems to be a need for tools suitable for selecting teaching methods that can improve students´ learning.

Our objective has been to create a concept inventory which can serve as a tool to assess different methods of teaching or interventions in the study of chemistry. Examples of such interventions can be to use videos as a supplement to text-books, laboratory activities or so called "inverted" or "flipped" classroom (Lage, Platt, & Treglia, 2000). The CCI may also be used for other purposes, such as studying connections between different misconceptions or mapping of which topics where the teaching should be reformed. Investigating connections between conceptions requires investigations beyond mapping scores of the complete inventory. This may nevertheless be achieved by comparing students answers to specific questions. Methods suited to measure learning efficacy can be used at universities for more than designing the chemistry study, for example to advice students about choosing the best study methods or influencing the content of chemistry teaching education. Other possible areas of use may be to test the knowledge developed during upper secondary chemistry classes. Concept inventory tests may also indicate quality differences in chemistry text-books, changes caused by altered curricula or other changes in the primary or secondary school system.

The developed chemistry concept inventory is designed to fit the current curricula in general chemistry courses at the Norwegian University of Science and Technology (NTNU). The curricula in these courses are to a high degree consistent with the content of widely used international editions of general chemistry textbooks for colleges and universities. Therefore, is seems probable that the inventory can be used at other teaching institutions, for similar purposes as those described in this article. The main aim has been oriented at general chemistry courses where the inventory will be used as a post-test. However, the test might also work as a tool for investigating students' individual learning by administering it both as a pre- and post-test. The development of alternative concept inventories can give similar tools, designed for other studies, curricula or narrowed towards one or a few chemistry topics instead of the whole curriculum. As indicated by Mazur (Mazur, 2009) and others, current assessment systems often fail to test students understanding of science concepts. The described concept inventory may also be used to adjust and improve student assessments. Analyses of test results in ongoing studies will give information about the present “state of the art” at our university, and also about the differences between students at this university and other teaching institutions.

5. Conclusions

A chemistry concept inventory for assessing students learning and identifying their alternative conceptions has been evaluated and tested. This inventory, CCI 3.0, aims at functioning as an assessment tool in chemistry education. Used as a post-test, it may give information about the effect of changes in teaching practices in university general chemistry courses. Used as a pre-test in first year university courses, the CCI 3.0 gives information about the learning outcome from upper secondary general science and chemistry courses. A concept inventory used for these purposes must be reliable and have necessary discriminatory power within the context where the inventory is administered. This inventory has been administered and evaluated using statistical tests, and results indicate that the concept inventory is a reliable and discriminating tool in the present context. The results of the tests can be grouped in the main knowledge areas to identify strengths and weaknesses in the students understanding, thus helping students to focus on the areas in need for improvement. The statistical analysis of the reliability and discriminatory power of CCI 3.0 shows that it can be applied within the context of general chemistry courses at NTNU and possibly other universities. The present form of the chemistry concept inventory is by no means the final, but will serve as a template for future versions as well as an inventory suitable for longitudinal studies. Such studies are started both at NTNU and at the university of Jyväskylä and may serve as a tool for investigating and comparing student's conceptions during their first year courses.

Access to the CCI may be given on request to corresponding author.

REFERENCES

Adams, W. K., Wieman, C. E., Perkins, K. K., & Barbara, J. (2008). Modifying and Validating the Colorado Learning Attitudes about Science Survey for Use in Chemistry. . *Journal of Chemical Education, 85*(10), s. 1435.

Bowen, D., & Bunce, D. M. (1997). Testing for Conceptual Understanding in General Chemisty 1. *The Chemical Educator, 2*, ss. 1-17.

Cloonan, C. A., & Hutchinson, S. (2010). A Chemistry Concept Reasoning Test. *Chemistry Education Research and Practice*, ss. 205-209.

Gabel, D. L., & Bunce, D. M. (1994). Research on problem solving: Chemistry. *Handbook of Research on Science Teaching and Learning, 11*, ss. 301-326.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *Phys. Teach., 30*, s. 141.

Kline, P. (1986). *A handbook of test conctruction: Introduction to psychometric design.* Methuen.

Krajcik, J. S. (1991). Developing Students' Understanding of Chemical Concepts. I S. M. Glynn, R. H. Yeany, B. K. Britton, & N. J. Lawrence Erlbaum (Red.), *The Psycology of Learning Science* (ss. 117-147).

Krause, S., Birk, J., Bauer, R., Jenkins, B., & Pavelich, M. J. (2004). Development, testing and application of a chemistry concept inventory. *Annual Frontiers in Education Conference.* Savannah.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*, ss. 151-160.

Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an invlusie learning environment. *The Journal of Economic Education, 31*(1), ss. 30-43.

Mayer, R. E. (2014). Principles for multimedia learning. California, USA: Harward University.

Mazur, E. (2009). Farewell, lecture. . *Science*, ss. 191-196.

Mulford, D. R., & Robinson, W. R. (2002). An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. *J. Chem. Educ*, *79*(6), s. 739.

Nakleh, M. B. (1992). Why som e students don't learn chemistry: Chemical misconceptions. *J. Chemi. Educ., 69*, ss. 191-196.

Persson, R. J. (2015). Evaluating the Force Concept Inventory for different student groups at the Norwegian University of Science and Technology. *arXiv preprint arXiv: 1504.06099*.

Pfundt, H., & Duit, R. (2000). *Bibliography. Students' alternative frameworks and science education.* Kiel, Germany: University of Kiel Institute for Science Education.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: a Constructivist Analysis of Knowledge in Transition. . *The Journal of the Learning Sciences*, *3*(2), ss. 115-163.

Stavy, R. (1991). Using analogy to overcome misconceptions about conservation of matter. *J. Res. Sci. Teach., 28*, ss. 305-313.

Stavy, R. (1995). Learning Science in the Schools. I L. Erlbaum:, *Research Informing Practice* (ss. 131-154). Hillsdale.

Taber, K. S. (2009). *Progressing Science Education: Constructing the scientific research programe into the contingent nature of learing science.* Dordrecht: Springer.

Treagust, D. F., & Chiu, M. (2011). Diagnostic assessment in chemistry. *Chemistry Education Research and Practice*, ss. 119-120.

van Marion, P. (2015). Praktisk arbeid. I P. v. (edt), *Biologididaktikk* (s. 257). 7094: Cappelen Damm Akademisk.

Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. . I D. Gabel (Red.), *Handbook of Research on Science Teaching and Learning.* New York: Macmillan.

THE OPPORTUNITIES AND CHALLENGES FOR ICT IN SCIENCE EDUCATION

Vesna Ferk Savec, University of Ljubljana, Faculty of Education, Slovenia

Received: 6.2.2017
Accepted: 7.8.2017

Abstract This article examines the opportunities and challenges for the use of ICT in science education in the light of science teachers' Technological Pedagogical Content Knowledge (TPACK). Some of the variables that have been studied with regard to the TPACK framework in science classrooms (such as teachers' self-efficacy, gender, teaching experience, teachers' beliefs, etc.) are reviewed, and variations of the TPACK framework specific for science education are expounded upon. In the conclusion, some of the aspects of TPACK in science education that need to be addressed in future are indicated, including the development of subject specific ICT-based resources and e-learning platforms; training to develop science teachers' integrated skills for the implementation of ICT in their subject teaching; the importance of the continuous encouraging of science teachers' for their participation in in-service training related to the use of ICT; and the examination of the role of science teachers' TPACK in the developing of students' 21st century trans-disciplinary skills.

Keywords science teachers, science education, ICT use in teaching, TPACK

1. Introduction

Since the beginning of the 21st century, with the expansion of the information age, societies have entered a process of change with the effect of rapidly advancing science and technology. For example, in the field of chemistry, the Chemical Abstracts Service (CAS), the world's authority for chemical information, registered the 100 millionth chemical substance in the CAS REGISTRY, in the 50th anniversary of the world's largest database of unique chemical substances. The number of new substances has been growing radically since the beginning of the century (Figure 1).

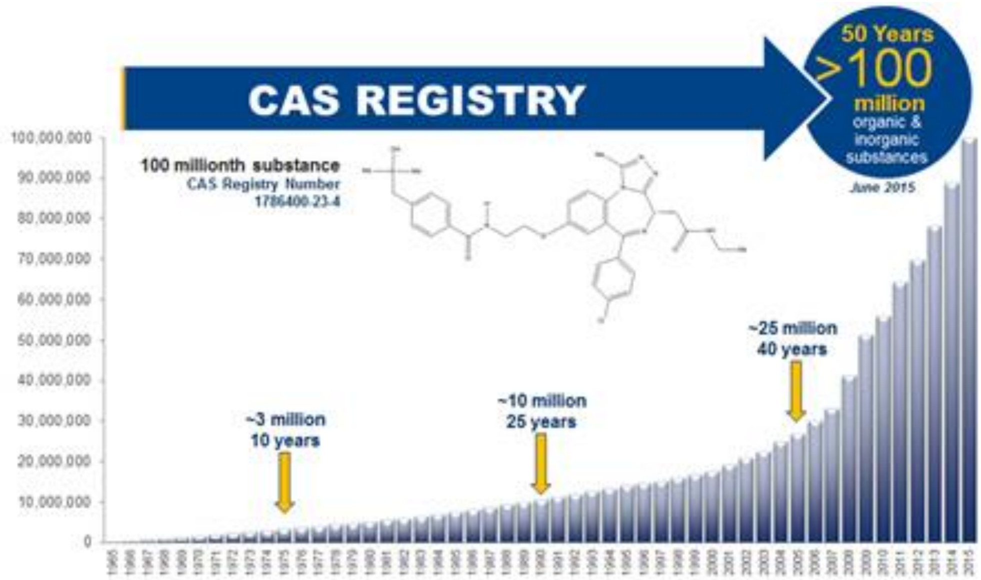


Figure 1: Growth in the CAS REGISTRY over 50 years (CAS, 2015)

In this process of change, information communication technology (ICT) tools resulting from new scientific and technological developments are present in every aspect of life, including education. Knowledge related to the effective use of educational technologies has become recognised as an important aspect of an educator's knowledge-base for the 21st Century (P21 Partnership for 21st Century Learning, 2016). In earlier stages of educational technology implementation, educators were taught in technology classes that focused primarily on technology skills independent from the pedagogical or content courses. As a result of educators' experience from school environment showing that technology skills alone are not enough (because one could know how to operate a piece of technology without knowing how to use it effectively to promote student learning), the focus then shifted to preparing educators to integrate technology into their teaching (Graham et al., 2009).

The rationale of the article and the research focus

In the information age, science teachers are continuously confronted with the occurrence of new opportunities and challenges, which are on one side the consequence of the rapid rate of discoveries in science and technology, and on the other side, the result of a remarkable development of information technology. Both simultaneously enable new possibilities and are a source of new ideas, that can be effectively implemented in teaching and learning processes in science education. This study recognises the answer for addressing contemporary opportunities and challenges in science education in the development of science teachers' Technological Pedagogical Content Knowledge (TPACK).

The research questions addressed in the article are:

1. What are up-to-date definitions of TPACK?
2. How is the TPACK framework implemented in science education?
3. What are open issues regarding TPACK in science education?

2. Technological Pedagogical Content Knowledge (TPACK)

In last decade, a significant amount of teaching and learning materials has been developed, which has been to a great extent influenced by recently emerged possibilities based on ICT, e.g. extensive use of visualisation such as simulations, animations, video, interactive learning environments, social media, augmented reality, etc. (Apotheker & Veenstra, 2015). However, it seems that in a significant portion of the teaching and learning materials, the focus has been on the implementation of new tools and quality, in the sense of learner's aspects, has been neglected (Vrtačnik & Ferk Savec, 2009). According to Tamis-LeMonda, Kuchirko, and Song (2014) and Hirsh-Pasek et al. (2015), when developing teaching and learning materials, it is crucial to consider that people learn best when they: (1) are actively involved ("minds-on"), (2) are engaged with the learning materials and undistracted by peripheral elements, (3) have meaningful experiences that relate to their lives, (4) socially interact with others in high-quality ways around new material, (5) within a context that provides clear learning goals.

To support the development of teachers' knowledge about the possibilities of effective integration of educational technology tools in classroom instruction of specific subjects, Mishra and Koehler (2006) developed the concept Technological Pedagogical Content Knowledge (TPACK), which builds on Shulman's (1987) pedagogical content knowledge (PCK) model. According to Koehler and Mishra (2008), TPACK is an integration of technological knowledge (TK - knowledge about technologies including the use of computers, the Internet, interactive whiteboard), content knowledge (CK - knowledge about the subject matter that is to be learned or taught), and pedagogical knowledge (PK - knowledge about the processes, practices or methods of teaching and learning) and is intended to help teachers to use technology effectively in their subject teaching (Figure 2).

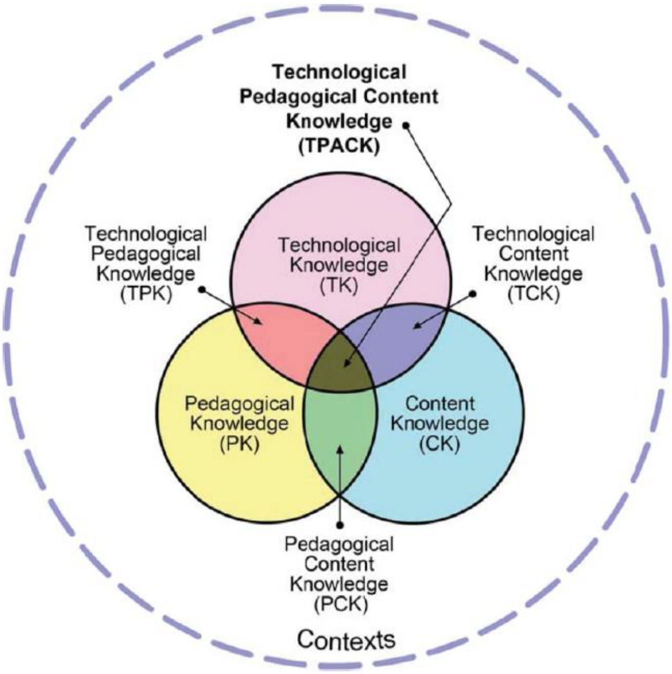


Figure 2: The seven components of TPACK (Koehler, 2016).

In other words, TPACK is a dynamic, integrative, and transformative knowledge of the technology, pedagogy, and content of a subject matter needed for pedagogically meaningful integration of ICT in teaching (Mishra & Kohler, 2006; Rogers & Twidle, 2014). Effective integration of technology and pedagogy around specific subject matter requires developing sensitivity to the dynamic relationship between these components of knowledge situated in unique contexts. Many factors, such as the personality of teachers, grade-level, school-specific factors, demographics, culture, etc., contribute that every context is unique, and no single combination of content, technology, and pedagogy will apply for every teacher, every course, or every perspective of teaching (Koehler & Mishra, 2009). The pillars and their integration into TPACK can be conceived as a continuum. The integrative view emphasises that teacher knowledge can be explained by the pillars per se, and TPACK is simply the sum of its parts. In contrast, the transformative view suggests that TPACK is a unique knowledge element that needs to be developed independently of its underlying constructs (Gess-Newsome 1999; Graham 2011). In a review of the TPACK research, Wu (2013) pointed that it has received increasing attention from researchers and educators during the past decade; in particular, the TPACK research increased at a fast pace from 2009. Regarding the distribution of the sample groups analysed by Wu (2013), the pre-service teachers' group has the highest ranking (54.2%), followed by the high school teachers' group (20.8%), the elementary school teachers' group (16.7%), and the university or college teachers' group (8%).

According to Porras-Hernandez and Salinas-Amescua (2013) and Rosenberg and Koehler (2015), the changes in teachers' TPACK are a function of micro factors at the classroom level (or learning environment), meso factors at the school level (or community level) and macro factors at the societal level. The changes in teachers' TPACK at all three levels is associated with teachers' factors and with students' factors. The proposed model presents the complexity and overlapping of factors that influence the development of teachers' TPACK.

3. Technological Pedagogical Content Knowledge in Science Education

Despite the growing and diverse research into many aspects of TPACK, it appears that the context remains an underdeveloped and under-researched component of the framework (Rosenberg & Koehler, 2015). Kelly (2010) examined whether the context was included in the conceptual definition of TPACK and found that it is frequently missing when researchers describe, explain, or operationalise TPACK in their work. In the review of research trends in TPACK, Wu (2013) found that more than half of the empirical TPACK studies focused on teachers' domain-general TPACK, and relatively fewer studies explored teachers' domain-specific TPACK. Science (20.8%) and mathematics (12.5%) were found to be the two major subject domains that were explored in those domain-specific TPACK studies. Wu (2013) suggested that this is probably because science and mathematics are relatively more abstract to students, and science teachers and math teachers may be more likely to adopt technologies to help students overcome their learning difficulties.

Based on the implementation of the TPACK framework in science classrooms, different variables have been examined, such as teachers' beliefs about self-efficacy (Lee & Tsai, 2008), skills of integrating technology into teaching (Guzey & Roehrig, 2009; Jang, 2010) and teachers' gender (Lin et al., 2012). Lin et al. (2012) found that female teachers were more confident in PK but less confident in TK in comparison to male teachers; however, in the study by Jang and Tsai (2013), male science teachers rated their technology knowledge significantly higher than female teachers did. Science teachers' TPACK was also found to relate to school context and their reasoning skills (Guzey & Roehrig, 2009). Research on science teachers' TPACK with regard to their teaching experience suggests varying results. Lee and Tsai (2008) found that more experienced teachers perceived their TPACK with respect to the Web lower than less experienced teachers did. In contrast, Jang and Tsai (2012) found that more experienced elementary science and mathematics teachers' TPACK were significantly higher than that of less experienced teachers. In the follow-up study, Jang and Tsai (2013) indicated that experienced science teachers tended to rate their content knowledge and pedagogical content knowledge in context (PCKCx) significantly higher than novice science teachers did. However, science teachers with less teaching experience tended to rate their technology knowledge and technological content knowledge in context (TPCKCx) significantly higher than teachers with more teaching experience did.

Helppolainen and Aksela (2015) examined chemistry teachers' knowledge, skills and beliefs on using ICT in education in comparison to other science teachers. They found that chemistry teachers' ICT knowledge, skills, beliefs, and usage were quite similar to those of other science teachers. Chemistry teachers reported to have good basic ICT skills, but they pointed to their lack of skills needed for ICT integration in chemistry teaching, indicating their deficiency of TCK or TPK. It was found that chemistry teachers had positive beliefs about the use of ICT in teaching and learning settings, but expressed the need for enough hardware and time, which might be the main reasons for a limited integration of ICT in their subject teaching, as well as hindering their development of their TPACK.

Recent articles have examined how chemistry teachers use Social Networks Sites (e.g. Facebook groups) and social media (e.g. YouTube) to facilitate learning. Blonder and Rap (2017) found that teachers' notion regarding what constitutes learning using chemistry Facebook groups had not changed during the teachers' training (workshop for chemistry teachers about technological tools, including Facebook groups), but the teachers' knowledge about how they can facilitate learning improved. Similarly, to improve the implementation of YouTube in chemistry teaching and learning, a one-year professional development course was designed to build the relevant TPCK for using videos in chemistry lessons, and to increase teachers' self-efficacy in editing and using videos in chemistry lessons (Blodnet et al., 2013). The research outcomes suggest that when the technology is readily available, such as YouTube videos, and the teachers receive the opportunity to develop skills, TPACK, and self-efficacy beliefs by experiencing the new technology in their own school practice by being a part of the community of learning, teachers will efficiently integrate the new technology in their teaching (Blonder et al., 2013).

Viewpoints studied regarding science teachers' domain-specific TPACK	**Selected bibliography**
(Student) teachers' perception, beliefs and attitudes towards TPACK	Lee & Tsai, 2008, Lin, Tsai, Chai, & Lee, 2012; Sancar-Tokmak, Surmeli, (Blonder et al., 2013); Ozgelen, 2014; Helppolainen & Aksela, 2015; Lehtinen, Nieminen, & Viiri, 2016; Blonder & Rap, 2017
Examining and developing of (student) teachers' TPACK	Guzey & Roehrig, 2009; Jang, 2010; (Jang & Chen, 2010); Jang & Tsai, 2013; Maeng, Mulvey, Smetana& Bell, 2013; (Blonder et al., 2013); Jaipal-Jamani & Figg, 2015, Blonder & Rap, 2017
Teachers' gender	Huang, & Fraser, 2009; Lin et al., 2012; Jang & Tsai, 2013; Chang, Tsai, & Jang, 2014
Teachers' teaching experience	Lee & Tsai, 2008; Jang & Tsai, 2012, 2013; Chang, Tsai, & Jang, 2014
Teachers' pedagogical reasoning skills	Guzey & Roehrig, 2009; Jaipal-Jamani & Figg, 2015

Table 1: Viewpoints addressed in examining science teachers' domain-specific TPACK

Some authors felt the need for the adaptation of TPACK for a specific subject domain. Jimonyiannis (2010a) developed the concept Technological Pedagogical Science Knowledge (TPASK), a framework for TPACK in science education. The TPASK model is based on three knowledge domains: pedagogical science knowledge (a science education operationalisation of PCK), technological science knowledge (a science education operationalisation of TCK), and TPK, but more oriented towards science education. According to Jimoyiannis (2010b, pp. 602), pedagogical science knowledge consists of several knowledge components e.g. scientific knowledge, science curriculum, transformation of scientific knowledge, students' learning difficulties about specific scientific fields, learning strategies, general pedagogy and educational context; technological science knowledge consists of e.g. resources and tools available for science subjects, operational and technical skills related to specific scientific knowledge, transformation of scientific knowledge, and transformation of scientific processes; technological pedagogical knowledge consists of e.g. ICT-based learning strategies, fostering scientific inquiry with ICT, supporting information skills, student scaffolding, and handling students' technical difficulties.

Another view of TPACK in science education is built on the idea of Magnusson, Krajcik, and Borko (1999) that science teachers' PCK is composed of (1) knowledge related to specific subject curricula, e.g. goals and objectives for specific contents in science curricula; (2) knowledge related to students' understanding of science, e.g. possible misconceptions regarding science concepts and process; (3) knowledge of instructional strategies, e.g. strategies and learning tools for specific science topics; and (4) knowledge of assessment of science knowledge and skills, e.g. methods of assessing experimental work in science learning. As reported by Yet et al. (2014), based on the competencies and knowledge that teachers are expected to develop for their PCK, a framework for TPACK-practical with eight knowledge dimensions in which science teachers practice teaching with technology has been proposed and validated: (1) Using ICT to understand students, (2) Using ICT to understand subject content, (3) Planning ICT infused curriculum, (4) Using ICT representations, (5) Using ICT integrated teaching strategies, (6) Applying ICT to instructional management, (7) Infusing ICT in teaching contexts, (8) Using ICT to assess students. The last dimension enables teachers to assess not only their students' learning progress but also the effectiveness of their instruction.

Lee and Tsai (2008) argued that with regard to teaching with the Internet, TPACK may be insufficient for providing adequate information to assist teacher preparation and professional development. They indicated that the Internet could be a specific form of technology, and introduced a framework of Technological Pedagogical Content Knowledge-Web (TPCK-W). Wang, Tsai, and Wei (2015) examined whether TPACK-W is a potential mediator contributing to the relationship between teaching and learning conceptions and science teaching self-efficacy by science teachers. They found that knowledge of and attitudes toward Internet-based instruction mediated the positive relationship between constructivist conceptions of teaching and learning and outcome expectancy, and that it also mediated the negative correlations between traditional conceptions of teaching and learning and science teaching efficacy.

4. Challenges related to Technological Pedagogical Content Knowledge (TPACK) in science education

Despite numerous studies dealing with various viewpoints regarding science teachers' domain-specific TPACK (Table 1) and the existence of well-elaborated modifications of the original TPACK framework for science education domain, in school practice many science teachers still do not use ICT in their lessons, and also there is insufficient evidence about how teachers, who claim to use ICT, implement it in their science classrooms.

The main challenge with regard to TPACK in science education, therefore, seems to be in finding ways to facilitate science teachers' recognition of the value of TPACK and how to facilitate their continuous care for the development of their own TPACK.

Niess, Sadri and Lee (2007) proposed a developmental model for TPACK in terms of teachers learning to integrate a technology into teaching. They found that teachers progressed through a five-stage developmental process when learning to integrate a particular technology in teaching and learning (Niess, 2011): (1) Recognising (knowledge), by which teachers are able to use the technology and recognise the alignment of the technology with subject matter content, yet do not integrate technology into the teaching and learning of the content; (2) Accepting (persuasion), by which teachers form a favourable or unfavourable attitude toward teaching and learning specific content topics with appropriate technology; (3) Adapting (decision), by which teachers engage in activities that lead to a choice to adopt or reject teaching and learning specific content topics with appropriate technology; (4) Exploring (implementation), by which teachers actively integrate teaching and learning of specific content topics with appropriate technology; (5) Advancing (confirmation), by which teachers redesign the curricula and evaluate the results of the decision to integrate teaching and learning specific content topics with appropriate technology.

An important caveat when considering these levels and the progression toward TPACK is that, while appearing linear with respect to a particular technology, the transition from one level to another does not display a regular, consistently increasing pattern. It requires rethinking the content and the pedagogies, thus, the levels are proposed to display more of an iterative process in the development of TPACK (Niess et al., 2009).Teachers often tend to use ICT largely to support, enhance and complement existing classroom practice rather than re-shaping subject content, goals, and pedagogies (Osborne & Hennessy, 2003). Regarding the state-of-the-art about the implementation of ICT in science teaching and learning, TIMSS (2015) reported considerable variation in computer availability among 57 participating countries for the use of computers in science lessons with fourth-graders, with an international average of 46%. It was found that, on average, more than one quarter of the fourth-grade students were asked by their science teachers to use computers at least monthly for activities, such as conducting scientific procedures or experiments (26%), practicing skills and procedures (31%), looking up ideas and information (41%) and studying natural phenomena through simulation (28%). For eighth-grade science, TIMSS (2015) also found considerable variation in computer availability among participating countries for the use in science lessons, with an international average of 42%. Thereby, 28-37% of eight-grade students were asked to use computers at least monthly, for various activities, such as conducting scientific procedures or experiments (28%), practicing skills and procedures (30%), looking up ideas and information (37%), studying natural phenomena through simulation (29%), and processing and analysing data (29%). It cannot be simply assumed that the introduction of ICT necessarily transforms science teaching and learning.

The critical role of the science teacher in creating the conditions for efficient ICT-supported learning through selecting and evaluating appropriate technological resources for achieving of selected curriculum aims, addressing students' needs, as well as in designing, structuring and sequencing the learning activities, needs to be acknowledged. In order to address the challenge of facilitating science teachers' continuous development of TPACK, it is important to efficiently overcome the constraints that might influence teachers' TPACK development. The possible constraints listed by science teachers' include lack of time to gain confidence and experience with ICT; limited accessibility to reliable resources; lack of time for critical selection of learning resources related to curriculum topics; a science curriculum overloaded with content; assessment that requires no use of the technology; the lack of subject-specific guidance for using ICT to support learning; lack of support for the installation and use of contemporary software in science classrooms; and lack of students' interest for learning science (Ferk Savec, 2015; Osborne & Hennessy, 2003). Most of the above-listed constraints are related to variables also recognised by Robinson (2003) and Inan & Lowther (2010) as the factors that affect technology integration by using path models to determine the relationships between teacher demographic characteristics (years of teaching and age), computer proficiency, external support variables (availability of computers, technical support, and overall support), teachers' perceptions (beliefs, readiness) toward using computers and usage of computers. Considering the factors that have been studied about their effects for integrating technology in teaching and learning, a visual description for thinking about factors that influence teachers' development of their TPACK is proposed in Figure 3.

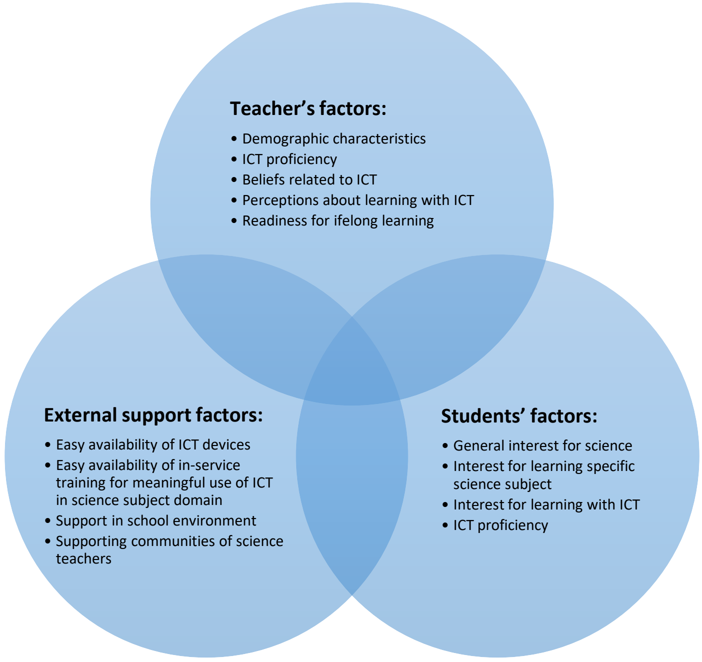


Figure 3: Factors that significantly influence on science teacher's development of TPACK.

5. Conclusions

Despite the rapid development of ICT and the great level of advances in science and technology in the previous decade, science teachers are facing many challenges and opportunities with regard to science education in their school practice. The article attempts to address some of the aspects relevant to this very current topic. Thereby, teachers' Technological Pedagogical Content knowledge (TPACK) can be recognised as a dynamic, integrative, and transformative knowledge of technology, pedagogy, and science contents needed for the pedagogically meaningful integration of ICT in science teaching (Mishra and Kohler, 2006; Rogers and Twidle, 2014). Although the general TPACK framework has been accepted in a wide range of areas, and also extensively used in science education field, some researchers have proposed variations of TPACK framework to address specific needs of science teachers; for example, Jimonyiannis (2010a) developed the concept of Technological Pedagogical Science Knowledge (TPASK), a framework for TPACK specific for science education.

In particular, the following issues need to be considered regarding science teachers' future development of TPACK:

(1) Continuous care for the availability of up-to-date ICT devices in science classrooms, including possibilities for the use of students' own devices;

(2) Availability of in-service training to develop and continuously update teachers' knowledge and skills for didactically meaningful implementation of ICT in the teaching practice of science subjects, although most science teachers already use ICT;

(2) Subject specific ICT-based resources and e-learning platforms accompanied by training need to be provided to teachers so that their technology related knowledge can be promoted;

(3) Encouragement of the participation of science teachers in ICT training to increase positive beliefs about teaching with ICT and understanding of its potential in improving students' learning outcomes in science;

(4) Organisation of training sessions for headmasters and school managers to support teachers in their continuous TPACK development;

(5) Supporting of the national and international communities of science teachers and their activities to support teachers' didactical use of ICT in teaching science subjects.

References

Apotheker, J. & Veldman, I. (2015). Twenty-first century skills: Using the web in chemistry education. In: J. Garcia-Martinez and E. Serrano-Torregrosa (Eds.) Chemistry education : best practices, opportunities and trends. Wiley-VCH, Weinheim, pp. 565-594. doi:10.1002/9783527679300.ch23

CAS (2015). CAS Assigns the 100 Millionth CAS Registry Number to a Substance Designed to Treat Acute Myeloid Leukemia. Retrieved from https://www.cas.org/news/media-releases/100-millionth-substance

Chang, Y., Tsai, M. F., & Jang, S. J. (2014). Exploring ICT use and TPACK of secondary science teachers in two contexts. US-China Education Review, 4(5), 298-311.

Blonder, R., Jonatan, M., Bar-Dov, Z., Benny, N., Rap, S., & Sakhnini, S. (2013). Can You Tube it? Providing chemistry teachers with technological tools and enhancing their self-efficacy beliefs. Chemistry Education Research and Practice, 14(3), 269-285. doi:10.1039/C3RP00001J

Inan, F. A., & Lowther, D. L. (2010). Factors affecting technology integration in K-12 classrooms: A path model. Educational Technology Research and Development, 58(2), 137-154.

Jang, S.-J., & Chen, K.-C. (2010). From PCK to TPACK: Developing a transformative model for pre-service science teachers. Journal of Science Education and Technology, 19(6), 553-564.

Niess, M. L. (2011). Investigating TPACK: Knowledge growth in teaching with technology. Journal of educational computing research, 44(3), 299-317.

Niess, M. L., Ronau, R. N., Shafer, K. G., Driskell, S. O., Harper, S. R., Johnston, C., . . . Kersaint, G. (2009). Mathematics teacher TPACK standards and development model. Contemporary Issues in Technology and Teacher Education, 9(1), 4-24.

Niess, M. L., Sadri, P., & Lee, K. (2007). Dynamic spreadsheets as learning technology tools: Developing teachers' technology pedagogical content knowledge (TPCK). American Educational Research Association.

Szeto, E., Cheng, A. Y.-N., & Hong, J.-C. (2016). Learning with Social Media: How do Preservice Teachers Integrate YouTube and Social Media in Teaching? The Asia-Pacific Education Researcher, 25(1), 35-44. doi:10.1007/s40299-015-0230-9

Franz, A. K. (2011). Organic chemistry YouTube writing assignment for large lecture classes. Journal of Chemical Education, 89(4), 497-501.

Ferk Savec, V. (2015). Priložnosti in izzivi za učitelje kemije v informacijski dobi. In: OREL, Mojca (Ed.). Modern approaches to teaching coming generation = Sodobni pristopi poučevanja prihajajočih generacij. Polhov Gradec: Eduvision, pp. 833-847.

Guzey, S. S. & Roehrig, G. H. (2009). Teaching science with technology: Case studies of science teachers' development of technology, pedagogy, and content knowledge. Contemporary Issues in Technology and Teacher Education, 9(1), 25-45.

Graham, C. R., Burgoyne, N., Cantrell, P., Smith, L., Clair, L. S., & Harris, R. (2009). TPACK development in science teaching: Measuring the TPACK confidence of inservice science teachers. (2009). TechTrends, 53(5), 70–79. doi:10.1007/s11528-009-0328-0

Gess-Newsome J .(1999). Pedagogical content knowledge: an introduction and orientation. In: Gess-Newsome J, Lederman N (Eds.): Nature, sources, and development of pedagogical content knowledge for science teaching. Kluwer, Dordrecht, pp. 3-17. doi:10.1007/0-306-47217-1_1

Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). Computers & Education, 57(3), 1953–1960. doi:10.1016/j.compedu.2011.04.010

Helppolainen, S. & Aksela, M. (2015). Science teachers' ICT use from a viewpoint of technological pedagogical content knowledge (TPCK). LUMAT, 3(6), 783-799.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting Education in "Educational" Apps: Lessons From the Science of Learning. Psychological Science in the Public Interest, 16(1), 3-34. doi:10.1177/1529100615569721

Huang, S. Y. L., & Fraser, B. J. (2009). Science teachers' perceptions of the school environment: Gender differences. Journal of Research in Science Teaching, 46(4), 404-420.

Jaipal-Jamani, K., & Figg, C. (2015). A case study of a TPACK-based approach to teacher professional development: Teaching science with blogs. Teacher Education, 15(2), 161-200.

Jang, S.-J. (2010). Integrating the interactive whiteboard and peer coaching to develop the TPACK of secondary science teachers. Computers & Education, 55(4), 1744-1751. doi:10.1016/j.compedu.2010.07.020

Jang, S.-J., & Tsai, M.-F. (2013). Exploring the TPACK of Taiwanese secondary school science teachers using a new contextualized TPACK model. Australasian Journal of Educational Technology, 29(4). doi:10.14742/ajet.282

Jang, S. J., & Chen, K. C. (2010). From PCK to TPACK: Developing a transformative model for pre-service science teachers. Journal of Science Education and Technology, 19(6), 553-564.

Jimoyiannis, A. (2010a). Designing and implementing an integrated technological pedagogical science knowledge framework for science teachers' professional development. Computers & Education, 55(3), 1259–1269. doi:10.1016/j.compedu.2010.05.022

Jimoyiannis, A. (2010b). Developing a technological pedagogical content knowledge framework for science education: implications of a teacher trainers' preparation program. Proceedings of Informing Science & IT Education Conference, pp. 597 – 607. Retrieved from http://proceedings.informingscience.org/InSITE2010/InSITE10p597-607Jimoyiannis867.pdf

Kelly, M. A. (2010). Technological Pedagogical Content Knowledge (TPACK): A Content analysis of 2006–2009 print journal articles. In D. Gibson, & B. Dodge (Eds.), Proceedings of the Society for Information Technology & Teacher Education. Chesapeake, VA: AACE, pp. 3880–3888.

Koehler, M. J. & Mishra, P. (2008). Introducing TPACK. In AACTE Committee on Innovation & Technology (Ed.), Handbook of technological pedagogical content knowledge for educators (pp. 3–29). NewYork: Routledge. doi:10.4324/9781315759630

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? Contemporary Issues in Technology and Teacher Education, 9(1), 60-70.

Koehler, M. (2016). TPACK Image. Retrieved from http://tpack.org; Reproduced by permission of the publisher, © 2012 by tpack.org

Lee, M.-H., & Tsai, C.-C. (2008). Exploring teachers' perceived self-efficacy and technological pedagogical content knowledge with respect to educational use of the World Wide Web. Instructional Science, 38(1), 1–21. doi:10.1007/s11251-008-9075-4

Lehtinen, A., Nieminen, P., & Viiri, J. (2016). Preservice teachers' TPACK beliefs and attitudes toward simulations. Contemporary Issues in Technology and Teacher Education, 16(2), 151-171.

Lichter, J. (2012). Using YouTube as a platform for teaching and learning solubility rules. Journal of Chemical Education, 89(9), 1133-1137.

Lin, T.-C., Tsai, C.-C., Chai, C. S., & Lee, M.-H. (2012). Identifying science teachers' perceptions of technological pedagogical and content knowledge (TPACK). Journal of Science Education and Technology, 22(3), 325–336. doi:10.1007/s10956-012-9396-6

Magnusson, S., Krajcik, J. & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. Lederman (Eds), PCK and science education. New York, NY: Kluwer Academic Publishers. pp. 3–17.

Maeng, J. L., Mulvey, B. K., Smetana, L. K., & Bell, R. L. (2013). Preservice teachers' TPACK: Using technology to support inquiry instruction. Journal of Science Education and Technology, 22(6), 838-857.

Mishra, P. & Koehler, M. J. (2006). Technological pedagogical content knowledge: a framework for integrating technology in teacher knowledge. Teachers College Record, 108(6), 1017–1054. doi:10.1111/j.1467-9620.2006.00684.x

Mishra, P., & Kereluik, K. (2011). What 21st century learning? A review and a synthesis. In M. Koehler & P. Mishra (Eds.), Proceedings of Society for Information Technology and Teacher Education International Conference 2011. Chesapeake, VA: AACE. pp. 3301–3312.

Koehler, M. J., Mishra, P. & Bouck, E. C., DeSchryver, M., Kereluik, K., Shin, T. S., & Wolf, L. G. (2011). Deep-play: developing TPACK for 21st century teachers. International Journal of Learning Technology, 6(2), 146. doi:10.1504/ijlt.2011.042646

Osborne, J. & Hennessy S. (2003). Literature review in science education and the role of ict: promise, problems and future directions. A NESTA Futurelab Research report. Retrieved from https://hal.archives-ouvertes.fr/file/index/docid/190441/filename/osborne-j-2003-r6.pdf

P21 Partnership for 21st century learning. (2016). Retrieved from http://www.p21.org/our-work/p21-framework

Porras-Hernandez, L. H. & Salinas-Amescua, B. (2013). Strengthening TPACK: A broader notion of context and the use of teacher's narratives to reveal knowledge construction. Journal of Educational Computing Research, 48(2), 223-244. doi: 10.2190/EC.48.2.f

Rogers, L., & Twidle, J. (2013). A pedagogical framework for developing innovative science teachers with ICT. Research in Science & Technological Education, 31(3), 227–251. doi:10.1080/02635143.2013.833900

Rosenberg, J. M., & Koehler, M. J. (2015). Context and technological pedagogical content knowledge (tpack): A systematic review. Journal of Research on Technology in Education, 47(3), 186–210. doi:10.1080/15391523.2015.1052663

Sancar-Tokmak, H., Surmeli, H., & Ozgelen, S. (2014). Preservice science teachers' perceptions of their tpack development after creating digital stories. International Journal of Environmental and Science Education, 9(3), 247-264.

Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. Harvard Educational Review, 57(1), 1–23. doi:10.17763/haer.57.1.j463w79r56455411

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? Current Directions in Psychological Science, 23(2), 121–126. doi:10.1177/0963721414522813

TIMSS (2015). IEA's Trends in International Mathematics and Science Study – TIMSS 2015. Retrieved from http://timss2015.org/download-center/

Vrtačnik, M. & Ferk Savec, V. (2009). Kako razvijati e-gradiva z dodano vrednostjo? = How could value-added e-units be developed?. In: OREL, Mojca (Ed.). Nova vizija tehnologij prihodnosti = The new vision of future technologies. Ljubljana: Evropska hiša, pp. 225-236.

Wang, Y.-L., Tsai, C.-C., & Wei, S.-H. (2015). The Sources of Science Teaching Self-efficacy among Elementary School Teachers: A mediational model approach. International Journal of Science Education, 37(14), 2264–2283. doi:10.1080/09500693.2015.1075077

Wu, Y.-T. (2013). Research trends in technological pedagogical content knowledge (TPACK) research: A review of empirical studies published in selected journals from 2002 to 2011. British Journal of Educational Technology, 44(3), E73–E76. doi:10.1111/j.1467-8535.2012.01349.x

Yeh, Y.-F., Hsu, Y.-S., Wu, H.-K., Hwang, F.-K., & Lin, T.-C. (2013). Developing and validating technological pedagogical content knowledge-practical (TPACK-Practical) through the Delphi survey technique. British Journal of Educational Technology, 45(4), 707–722. doi:10.1111/bjet.12078

KOLMETOISTAVUOTIAIDEN NUORTEN KÄSITYKSIÄ LUONNONTIETEELLISESTÄ TUTKIMUKSESTA

Arvi Hakanen ja Jari Lavonen
University of Helsinki, Department of education

Received: 27.12.2016
Accepted: 10.8.2017

Tiivistelmä Luonnontieteellisen tutkimuksen luonteen ymmärtämistä mittaavaa VASI-testiä (views about the scientific inquiry) sovellettiin 149:n seitsemäsluokkalaisen muodostamaan otokseen. Vastaukset luokiteltiin neljään ryhmään (vastaus puuttuu, vastaus ei ole mallitiedon mukainen, vastaus on osittain mallitiedon mukainen, vastaus on mallitiedon mukainen). Tutkimuksen perusteella läheskään kaikkia luonnontieteellisen tutkimuksen luonteen piirteitä ei hallita koherentisti seitsemännellä luokalla. Kuitenkin esimerkiksi VASI-näkökulma ”aineisto on eri asia kuin tulokset” osattiin ilman eksplisiittistä opetusta ja todennäköisesti siksi, että suomalaisilla oppilailla on keskimäärin hyvät valmiudet loogiseen päättelyyn ja riittävät kielelliset valmiudet arvioida käsitteiden välisiä eroja/yhtäläisyyksiä.

7TH GRADE STUDENTS' VIEWS OF THE SCIENTIFIC INQUIRY

Abstract VASI (views about the scientific inquiry) questionnaire was applied to a sample of 149 7th grade students. When assessing each aspect of VASI, pupils' views were categorized into one of four categories: informed, mixed, naive, and unclear. According to the diagnostical test, 7th grade students did not possess informed conceptions of NOS in average, while certain questions were better understood. E. g., question #4 which measured the possession of the VASI aspect "Data does not equal evidence" was relatively well understood indicating that Finnish students had good reading ability and logical deduction skills without prior explicit teaching of VASI concepts.

1. Johdanto

Käsite “luonnontieteiden luonne” (NOS, nature of science) on esiintynyt anglosaksisissa maissa opetussuunnitelmateksteissä yli 100 vuotta, ja sitä koskevaa kasvatustieteellistä tutkimusta on tehty yli 50 vuotta (Lederman, 2007). Keskeistä luonnontieteiden luonteen ymmärtämisessä on luonnontieteellisen tutkimuksen rakenteen ja vaiheiden sekä luonnontieteellisten selitysten luonteen ymmärtäminen. Nämä on otettu myös PISA 2006 viitekehyksessä keskeisiksi NOS-osaamisen alueiksi (OECD, 2007).

Luonnontieteiden luonteen oppimisen tavoittelu implisiittisesti luonnontieteellisten kokeiden ohessa tulee hyvin esille siinä, millä tavalla opetussuunnitelman perusteasiakirjat (OPS) kuvaavat kokeelliselle työskentelylle asetettuja tavoitteita. Vuoden 1994 OPS (Opetushallitus, 1994) kuvaa asiaa siten, että kokeellisen työskentelyn on tarkoitus tukea lähinnä käsitteiden oppimista. Vuoden 2004 OPS:ssa todetaan (Opetushallitus, 2004): ”Kokeellisuuden tehtävänä on auttaa oppilasta hahmottamaan luonnontieteiden luonnetta ja omaksumaan uusia luonnontieteellisiä käsitteitä, periaatteita ja malleja sekä kehittää kokeellisen työskentelyn ja yhteistyön taitoja ja innostaa oppilasta fysiikan/ kemian opiskeluun”. Vastaavasti vuoden 2014 OPS:ssa todetaan (Opetushallitus, 2014): ”... opetuksen lähtökohtana ovat luonnosta ja teknologisesta ympäristöstä tehdyt havainnot ja tutkimukset. Kokeellisuudella on oleellinen merkitys käsitteiden omaksumisessa ja ymmärtämisessä, tutkimisen taitojen oppimisessa ja luonnontieteiden luonteen hahmottamisessa. ...”

OPS:ssa luonnontieteen luonteen hahmottaminen mainitaan tavoitteena fysiikan ja kemian opetuksen tavoitteiden joukossa ensimmäisen kerran vuonna 2004. Vuoden 2014 OPS:ssa kuvataan lisäksi opetuksen tavoitteisiin liittyvät päättöarvioinnin kriteerit hyvälle osaamiselle peruskoulun päättövaiheessa. Fysiikassa ja kemiassa opetuksen tavoitteissa ja arviointikriteereissä luonnontieteiden luonteen hahmottaminen mainitaan eksplisiittisesti yhtenä luonnontieteiden osaamisalueena. Fysiikan ja kemian päättöarvioinnin kriteereissä luonnontieteellisen tiedon luonne mainitaan kolmeen kertaan taulukon 1 mukaisesti. Myös biologiassa ja maantiedossa luonnontieteen luonne sisältyy implisiittisesti biologian ja maantieteen opetuksen taitotavoitteisiin, joissa mainitaan mm. “luonnontieteellinen/ maantieteellinen ajattelutaito”.

Taulukko 1. Luonnontieteiden luonteen osaamisen kriteerit osana fysiikan ja kemian päättöarvioinnin kriteereitä perusopetuksen opetussuunnitelman perusteissa (Opetushallitus, 2014).

Opetussuunnitelman kohta	Kuvaus
Opetuksen tavoite	T13 Ohjata oppilasta hahmottamaan luonnontieteellisen tiedon luonnetta ja kehittymistä sekä tieteellisiä tapoja tuottaa tietoa
Sisältöalueet	S1 Luonnontieteellinen tutkimus S4 Fysiikka/kemia maailmankuvan rakentajana
Arvioinnin kohde	Luonnontieteellisen tiedon luonteen hahmottaminen
Hyvä osaaminen	Oppilas osaa kuvata fysiikkaan/kemiaan liittyvien esimerkkien avulla luonnontieteellisen tiedon luonnetta ja kehittymistä. Oppilas osaa kuvata esimerkkien avulla tieteellisiä tapoja tuottaa tietoa.

Oppilaan ja opettajan NOS-käsityksiä on tutkittu lukuisissa tutkimuksissa, sekä erikseen että yhdessä, kirjallisin kyselyin ja haastatteluin sekä niiden erilaisin variaatioin ja yhdistelmin (Lederman, 1992, 2007). Käsityksellä tarkoitetaan tässä kokemusten ja ajattelun avulla saatua kuvaa ilmiöstä, tässä luonnontieteiden luonteesta (Ahonen, 1994). Havainnot aikaisemmista oppilaiden käsityksiä luonnontieteiden luonteesta kartoittaneista tutkimuksista voidaan tiivistää seuraavasti (Lederman, 2007):

i. Kolmetoistavuotiaat eivät tyypillisesti osaa antaa informatiivisia ja asiallisia vastauksia kysymyksiin, jotka kartoittavat heidän käsityksiään luonnontieteiden luonteesta,
ii. Luonnontieteiden luonnetta opitaan ymmärtämään eksplisiittisesti opettamalla paremmin kuin implisiittisesti esimerkiksi oppilastöiden yhteydessä,
iii. Opettajan luonnontieteiden luonteen ymmärtäminen ei siirry automaattisesti osaksi opetuskäytäntöjä,
iv. Luonnontieteiden luonteen oppimista ei tyypillisesti pidetä yhtä arvokkaana tavoitteena kuin luonnonlakien ja luonnontieteellisten teorioiden oppimista.

Kohdan (ii) mukaan implisiittinen luonnontieteiden luonteen opetus, jossa luonnontieteiden oppiminen kytkeytyy löyhästi opiskelutilanteisiin, ei johda luonnontieteiden luonteen monipuoliseen hahmottamiseen. Khishfe ja Abd-El-Khalick (2002) väittävät perustellen, että vain eksplisiittinen NOS-opetus voi johtaa luonnontieteiden luonteen oppimiseen. Väite voidaan kyseenalaistaa Suomessa, sillä luonnontieteiden luonteen opetus on mainittu vasta vuoden 2004 OPS:ssa tavoitteena ja sitä ennen se on siis ollut lähinnä implisiittistä (kokeisiin ja laboratoriotöihin integroitua). Tästä huolimatta suomalaiset oppilaat ovat saavuttaneet hyviä tuloksia PISA-tutkimusten luonnontieteiden kaikissa osioissa (Lavonen & Laaksonen, 2009). Vuoden 2015 PISA-tutkimuksessa (OECD, 2016) suomalaiset nuoret saivat luonnontieteellisen sisältötiedon osaamisessa 534 pistettä ja menetelmällisen ja episteemisen tiedon osaamisessa 528 pistettä (OECD-maiden keskiarvona saatava vertailutaso oli kummassakin 493 pistettä). Suomalaiset tutkijat ovat myös esittäneet, ettei ole riittävää puhua NOS-käsitteistä yleisinä (dekontekstualisoituina) oppisisältöinä, vaan tarvitaan eri tieteenaloihin ja oppiaineisiin kontekstualisoituja NOS-sisältöjä ja tieteenfilosofista, kasvatustieteellistä, historiallista ja yhteiskunnallista keskustelua näiden opettamisesta ja oppimisesta (Tala & Vesterinen, 2015).

Tieteenfilosofisesti orientoituneessa tutkimuksessa on pyritty hahmottamaan tulkintaa siitä, mitä luonnontieteiden luonteesta olisi opittava ja opetettava. Opetussuunnitelmien konteksteissa käytettävään luonnontieteiden luonteen konsensustulkintaan on tyypillisesti sisällytetty seuraavat piirteet (Lederman, 2007):

> ”Luonnontieteellinen tieto on muuttuvaa, kokeellista ja subjektiivista, se sisältää inhimillistä luovuutta ja mielikuvitusta, se hyväksytään sosiaalisissa suhteissa ja sen hyväksyntä riippuu kulttuurista. Kaikkia ilmiöiden välisiä riippuvuuksia ei voida aistein havaita, ja laki ja teoria ovat eri asioita (teoria on joukko yhteen kuuluvia lakeja).”

Tulkinnassa ei kerrota suoraan mitä kokeellisella tarkoitetaan tai mitä empiirisestä metodista ja tutkimuksen tekemisestä tulisi oppia ja opettaa. Yhtenä vaihtoehtona on laatia lista tieteellisen tutkimuksen osista ja vaiheista ja katsoa millä logiikalla tieteellisessä tutkimuksessa edetään, ts. miten tutkimuksen osat liittyvät toisiinsa ja miten eri vaiheet muodostavat loogisesti etenevän kokonaisuuden. Tämä on keskeistä käsillä olevassa VASI-tutkimuksessa, jossa nk. VASI-näkökulmat täydentävät em. tulkintaa. VASI on jatkoa aikaisemmille VNOS- ja VOSI- tutkimuksille, joissa tavoitteena on ollut oppilaan luonnontieteiden luonnetta ja tieteellistä tutkimusta koskevien käsitysten kartoittaminen (assessment) (Lederman & Lederman, 2014).

Luonnontieteissä ja tieteessä ylipäätään on problematisoitu teorian ja käytännön suhdetta samoin kuin sitä onko suositeltavampaa edetä yksityistapauksista yleistyksiin (induktio) vai päinvastoin (deduktio). Tieteenalojen jakautuminen teoreettiseen ja käytännölliseen (fysiikkaan, filosofiaan) samoin kuin koulussa opetettavat monet tiedeaineet ovat seurausta hyvin pitkäaikaisista ”universaalikiistoista”. Lederman ja Lederman (2014) väittävät, että useat tiedemaailmaa jakavat universaalikiistat eivät ole relevantteja yläkoulun kontekstissa. Sitä vastoin kysymys siitä, onko evoluutioteoria teoria vai laki, on relevantti kysymys yläkoulussa, jossa opiskellaan yleisiä luonnontieteiden luonteeseen kuuluvia periaatteita tuottaa, arvioida ja perustella tietoa, muttei tieteessä, jossa nämä periaatteet jo hallitaan.

2. Tutkimuksen toteutus

2.1. Tutkimuskysymys

Tässä työssä raportoitu tutkimus on osa professori N. Ledermanin johtamaa kansainvälistä VASI (views about the scientific inquiry) -tutkimusta. VASI-tutkimuksessa tarkastellaan 13-vuotiaiden oppilaiden luonnontieteelliseen tutkimukseen liittyviä käsityksiä. Tutkimuksen kohteena on siten luonnontieteiden luonteeseen (NOS) liittyvä tieteellistä tutkimusta (SI) koskevien käsitysten osa. Tutkimuksen lähtökohtana on havainto, jonka mukaan opetussuunnitelman ja oppimateriaalien laatijat ja opettajat näyttäisivät yhä hiukan kevyesti olettavan, että luonnontieteellisen tutkimuksen luonne hahmottuisi oppilaalle pienimuotoisia tutkimuksia tekemällä ja kokeellisuuden ohessa ilman luonnontieteiden luonteeseen liittyvien peruskäsitteiden eksplisiittistä opetusta (Lederman, Lederman, Bartos, Bartels, Meyer & Schwartz, 2014).

VASIssa tavoitteena on hankkia tietoa siitä, mitä oppilaat osaavat luonnontieteellisen tutkimuksen luonteesta ja mitä oppilaille tulisi siitä opettaa. Luonnontieteellisen tutkimuksen luonteen ymmärtämistä mitataan oppilaiden käsityksiä kartoittavalla diagnostisella testillä. Sen sijaan, että oppilailta kysyttäisiin suoraan heidän näkemyksiään tieteellisestä tutkimuksesta, heidät pannaan testiin, jossa he joutuvat analysoimaan tutkimuksellisia tilanteita käyttäen omia tietojaan ja näkemyksiään hyväksi. Osassa tehtävistä näkemyksiä kysytään suoraan ja osassa epäsuoraan, esittämällä tieteellistä tutkimusta koskevia väittämiä ja pyytämällä perustellen kertomaan onko asia niin vai näin. Testivastauksia luokitellaan kolmeen kategoriaan sen mukaan ovatko oppilaan antamat vastaukset yhteensopivia/ jossain määrin/ ei lainkaan yhteensopivia niiden tieteellistä tutkimusta koskevien väittämien kanssa, jotka muodostavat testin taustalla olevan mallitiedon (laajan tutkijajoukon muodostama konsensus tieteellistä tutkimusta koskevista väittämistä).

Tavoitteena on selvittää, millainen käsitys, ts. opiskelun, kokemusten ja ajattelun avulla muodostunut kuva, peruskoulun seitsemäsluokkalaisille on hahmottunut luonnontieteellisestä tutkimuksesta siinä vaiheessa, kun fysiikkaa ja kemiaa on opiskeltu puoli vuotta aineenopettajan johdolla. Opinnäytetyön tutkimuskysymykseksi muotoiltiin: Kuinka hyvin kolmetoistavuotiaat tytöt ja pojat analysoivat esitettyjä luonnontieteellisen tutkimuksen tilanteita tukeutumalla omiin tietoihinsa luonnontieteiden luonteesta? VASI-tutkimuksen Suomen osuus toteutettiin lomakekyselynä. Aineiston analyysissa voitiin vertailla eri käsitysten osaamista ja tarkastella tyttöjen ja poikien osaamisen eroja.

Osaamista mittaavissa tutkimuksissa tarkastellaan usein erilaisilla periaatteilla muodostettujen ryhmien osaamisessa olevia eroja, jotta voitaisiin päätellä onko osaaminen samankaltaista erilaisissa ryhmissä. Vertailuja tehdään, jotta voidaan päätellä tavoittaako opetus samalla tavalla erilaiset ryhmät. Esimerkiksi PISA-tutkimuksessa vertaillaan osaamista ryhmien välillä, jotka on muodostettu sukupuolen, maantieteellisen sijainnin, vanhempien sosio-ekonomisen taustan, kotikielen ja oppilaiden menestymisen perusteella (Simon, 2000). Tässä tutkimuksessa on mielekästä vertailla ryhmiä, jotka on muodostettu sukupuolen mukaan, sillä oppilaiden osaamisessa ja kiinnostumisessa on havaittu olevan eroja mitattaessa osaamista tai kiinnostumista sitä varten laaditun mittarin avulla. Havaitut erot voivat aiheutua osaamisessa tai kiinnostumisessa olevista eroista. Esimerkiksi vuoden 2015 PISA-tutkimuksen mukaan suomalaiset tytöt osaavat luonnontieteitä poikia merkitsevästi paremmin (OECD, 2016).

Sukupuolella tarkoitetaan yleensä kahta eri asiaa: biologista ('sex') ja sosiaalista ('gender') sukupuolta. Sosiaalinen sukupuoli rakentuu sosiaalisisessa vuorovaikutuksissa ja kulttuurin tukemana (Palmu, 2003). Luonnontieteiden tunneilla sosiaalisen sukupuolen muotoutumiseen vaikuttaa se, millä tavalla eri sukupuolta olevia henkilöitä ohjataan toimimaan tunneilla ja millä tavalla tunneilla käsiteltävissä esimerkeissä eri sukupuolta olevat henkilöt toimivat tai ajattelevat. Yleisesti ajatellaan, että toiminta ja esimerkit vaikuttavat siihen, että pojat ovat tyttöjä kiinnostuneempia luonnontieteistä, erityisesti fysiikasta, ja teknologiasta (Fairbrother, 2000). Kiinnostus vaikuttaa puolestaan siihen, millä tavalla opiskeltavia asioita opiskellaan. Koska PISAssa on havaittu tyttöjen osaavan luonnontieteitä poikia paremmin, on mielekästä tarkastella tyttöjen ja poikien osaamisen eroja myös luonnontieteellisen tutkimuksen luonteen ymmärtämistä mittaavassa VASIssa.

2.2. Kyselylomake

VASI-kysely on jatkoa aikaisemmille VNOS- ja VOSI-kyselytutkimuksille (Schwartz, Lederman & Lederman, 2008). Kysymyslomakkeen laadinnan pohjana olevat nk. VASI-näkökulmat ovat samoja, joita käytetään kahdessa yhdysvaltalaisessa opetussuunnitelmien perusteasiakirjassa (National Research Council, 2011 ja Achieve, Inc., 2013). VASI-testin tehtäväkohdat mittaavat monipuolisesti sekä teoreettisia että episteemisiä ja menetelmällisiä tutkimuksen teon taitoja. Ledermanin ym. integroiva lähestyminen luonnontieteiden luonteen ja luonnontieteellisen tutkimuksen teemaan käy ilmi ryhmän viimeaikaisista julkaisuista (Lederman, Lederman & Antink, 2013, Lederman & Lederman, 2014).

VASI-näkökulmia on yhteensä kahdeksan kappaletta, jotka on luokiteltu kirjaimiin A-H. Nämä näkökulmat näkyvät taulukossa 2. Taulukkoon 2 on myös jokaisen näkökulman kohdalle laitettu kyselylomakkeen tehtävä tai tehtäväkohta (1a-7b), jonka kautta näkökulmaa on tarkasteltu. Vastaamisaikaa taulukossa 2 esitettyihin seitsemään tehtävään oli 45 minuuttia.

Taulukko 2. VASI-näkökulmat ja niihin liittyvät kyselyn tehtäväkohdat.

A. Tutkimus alkaa tutkimuskysymyksellä

2) Kahdelta oppilaalta kysyttiin alkaako luonnontieteellinen tutkimus aina tutkimuskysymyksellä? Toinen oppilaista vastasi “kyllä” ja toinen “ei”. Kumman kanssa sinä olet samaa mieltä ja miksi?

B. Ei ole yhtä oikeaa tutkimusmenetelmää

1) Linnuista innostunut henkilö tarkkaili satoja erilaisia lintuja, jotka söivät erilaista ruokaa. Hän havaitsi, että samantapaista ruokaa syövillä linnuilla oli samantapainen nokka. Esimerkiksi linnuilla, jotka söivät kovia siemeniä, oli lyhyt ja voimakas nokka; ja linnuilla, jotka söivät hyönteisiä, oli ohut ja pitkä nokka. Hän rupesi ihmettelemään voisiko nokan muodolla ja ruualla, jota lintu syö, olla yhteys. Hän alkoi kerätä aineistoa (havaintoja), jotta voisi vastata kysymykseen. Hän päätteli kerätyn aineiston avulla, että linnun nokan muodon ja ruuan välillä on yhteys.

a. Pidätkö henkilön tekemää tutkimusta tieteellisenä? Ole hyvä ja perustele vastauksesi.
b. Onko henkilön tekemä tutkimus mielestäsi luonnontieteellinen koe? Ole hyvä ja perustele vastauksesi.
c. Voidaanko luonnontieteellisessä tutkimuksessa mielestäsi käyttää useampaa kuin yhtä menetelmää?

Jos vastasit c-kohtaan ei, kerro miksi luonnontieteellisen tutkimuksen tekemisessä on vain yksi menetelmä? Jos vastasit c-kohtaan kyllä, kuvaile kahta tutkimusta, joissa käytetään eri menetelmiä. Kerro, millä tavalla menetelmät poikkeavat toisistaan ja millä perusteella niitä voidaan pitää silti tieteellisinä.

C. Kysymys ohjaa menetelmän valintaa

5) Kaksi tutkijaryhmää käveli tutkimuslaboratorioihinsa. Tutkijat näkivät tien reunassa auton, jonka rengas oli tyhjä. Ryhmät ihmettelivät “menevätkö tietyn tyyppiset renkaat helpommin rikki?”

Ryhmä A meni laboratorioon ja alkoi tutkia erilaisten renkaiden käyttäytymistä samanlaisella tienpinnalla.

Ryhmä B meni laboratorioon ja alkoi tutkia samanlaisten renkaiden käyttäytymistä kolmella erilaisella tienpinnalla. Selitä miksi toisen ryhmän tutkimusmenetelmä on parempi kuin toisen.

D. Sama menetelmä ei takaa samaa tulosta

3) a. Jos useampi tutkija esittää saman tutkimuskysymyksen ja käyttää samaa menetelmää aineiston kokoamisessa, päätyvätkö he samoihin johtopäätöksiin? Kerro miksi tai miksi ei.

E. Menetelmän valinta voi vaikuttaa tuloksiin

3) b. Jos useampi tutkija esittää saman tutkimuskysymyksen ja käyttää eri menetelmää aineiston kokoamisessa, päätyvätkö he samoihin johtopäätöksiin? Kerro miksi tai miksi ei.

F. Johtopäätökset ovat linjassa aineiston kanssa

6) Taulukossa on esitetty kasvin viikoittaisen pituuskasvun yhteys kasvin saamaan valoon.

Valoisan ajan pituus minuutteina kunakin päivänä	**Kasvin pituuskasvu (cm viikossa)**
0	25
5	20
10	15
15	5
20	10
25	0

Aineistoon tukeutuen kerro mikä seuraavista johtopäätöksistä on mielestäsi oikea ja miksi.
Rengasta yksi vaihtoehto:

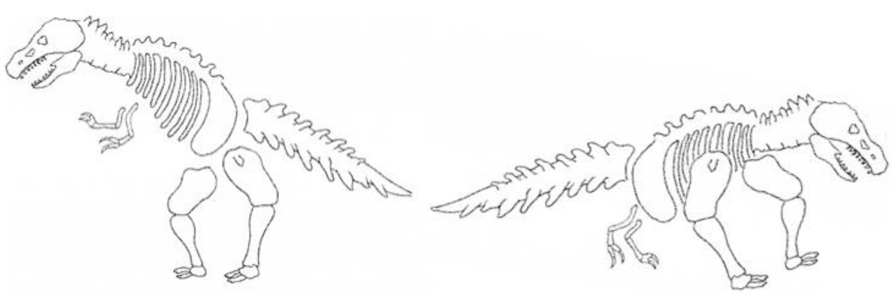
a. Kasvit kasvavat sitä pidemmiksi, mitä pidempi on valoisa aika.
b. Kasvit kasvavat sitä pidemmiksi, mitä lyhyempi on valoisa aika.
c. Kasvien pituuskasvun ja valoisan ajan pituuden välisestä yhteydestä ei voida päätellä mitään.

G. Johtopäätökset vs. aineisto

4) Ole hyvä ja selitä ovatko “aineisto” ja “tulokset” sama asia vai eri asia.

H. Johtopäätöksissä yhdistyvät aineisto ja tieto asioista

7) Ryhmä tutkijoita on löytänyt dinosaurusfossiilin luita. Tutkijat esittivät kaksi mahdollista rakennetta dinosauruksen luurangolle.



Kuva 1. Dinosaurusfossiilin mahdolliset rakenteet: vasemmalla on kuva 1a ja oikealla kuva 1b.

a. Esitä vähintään kaksi syytä sille, miksi useimmat tutkijat kannattavat kuvassa 1a esitettyä luurangon rakennetta.
b. Tarkastele vastaustasi yllä. Mitä informaatiota tutkijat käyttivät selittääkseen johtopäätöksensä?

2.3. Kyselyn käännökset ja takaisinkäännökset

Englanninkielinen kysely käännettiin suomeksi siten, että käännöksessä otettiin huomioon kielten väliset ominaispiirteet ja erot. Käännöksen validius varmistettiin siten, että suomenkieliset kysymykset käännettiin takaisin englanniksi ja takaisinkäännökset lähetettiin Yhdysvaltoihin arvioitaviksi. Joistakin takaisinkäännöksistä keskusteltiin ja etsittiin yhdessä sellaisia luontevia ilmaisuja, että ilmaisujen merkitykset säilyivät samoina. Seuraavat kaksi esimerkkiä havainnollistavat käännösprosessin läpikäyneitä ja hyväksyttyjä takaisinkäännöksiä:

1) Alkuperäinen: He began to collect data to answer the question…
   Takaisinkäännös: He started observations in order to answer the question…

2) Alkuperäinen: Given this data, explain which one of the following conclusions you agree and why.
   Takaisinkäännös: Rely on this data and explain which of the items you agree and why.

Kääntämisen kannalta käsitteet “tulokset” ja “johtopäätökset” ovat lähellä toisiaan, ja valinta näiden välillä tehtiin suomenkielisten kysymysten ymmärrettävyyden ja selkeyden perusteella. Asiayhteydestä riippuen “data” käännettiin “aineistoksi” tai "havainnoiksi".

2.4. VASI-arviointi, arviointikriteerit

Taulukko 3. VASI-kyselyn arviontiasteikko on neliportainen järjestysasteikko.

Vastausluokka	Koodi
C-vastaus on sellainen, joka on yhteensopiva asianomaisen VASI-näkökulman kanssa	3
M-vastaus sisältää sekä C-vastauksen että N-vastauksen piirteitä	2
N-vastaus on sellainen, joka ei ole yhteensopiva asianomaisen VASI-näkökulman kanssa	1
U-vastaus on epäselvä tai puuttuu kokonaan	0

Neliportainen asteikko koodattiin siis numeroiksi 0-3, jolloin vastausdataa voitiin analysoida, laskea mediaanit ja piirtää tulosten jakaumat.

Kyselyn alkutekstissä oppilasta kannustettiin vastaamaan rohkeasti kertomalla, että tehtäviin ei ole olemassa yhtä oikeaa vastausta. Vastausten arvioinnissa sovelletaan kuitenkin taulukon 2 mukaisesti VASI-näkökulmia, joiden kanssa oppilaan antaman asiallisen vastauksen tulee olla yhteensopiva, esim. tehtävässä 1 luonnontieteellisessä tutkimuksessa voidaan käyttää useampaa kuin yhtä menetelmää, tehtävässä 2 luonnontieteellinen tutkimus alkaa tutkimuskysymyksellä, tehtävässä 3a sama menetelmä ei takaa samaa johtopäätöstä, tehtävässä 3b eri menetelmät voivat johtaa erilaisiin johtopäätöksiin ja tehtävässä 4 aineisto on eri asia kuin tulokset. Tehtävään 7a annettavan vastauksen tulee sisältää kaksi syytä sille, miksi useimmat tutkijat kannattavat kuvassa 1a esitettyä dinosaurusfossiilin luurangon rakennetta (oikea rakenne, kuvassa 1b fossiilin etu- ja takaraajojen luut olivat vaihtaneet paikkaa), ja tehtävän 7b vastauksen, että johtopäätös edellyttää havaintoa ja havainnon alaan kuuluvaa tietoa asioista.

Seuraavan kolmen kappaleen avulla valotetaan vastausten arviointiprosessia ja annetaan esimerkkejä oppilasvastauksista tehtäviin 1a-1c.

Tehtävässä 1b kysytään ”Onko (tehtävän johdannossa kuvatun) henkilön tekemä tutkimus mielestäsi luonnontieteellinen koe?” Charles Darwinin tutkimaa Galapagossaarten peippojen evoluutiota 500 km:n etäisyydellä mantereesta voidaan pitää luonnon järjestämänä kokeena. Koeasetelman suunnittelua ja empiirisen aineiston hankkimista voidaan pitää luonnontieteellisen tutkimuksen riittävänä muttei välttämättömänä edellytyksenä. Varsinaisten kokeiden järjestämisen ja suunnittelun ohella luonnontieteellisen tutkimuksen teossa voi olla käytössä muitakin menetelmiä kuten havainnointia, tarkkailua ja seurantaa. Oppilasvastaukset olivat enimmäkseen n-vastauksia, esim. ”Tutkimus oli luonnontieteellinen koe, koska siinä tutkittiin lintuja.” Esimerkki m-vastauksesta: "On, koska linnut ovat luonnostaan syntyneet näin ja tämä henkilö on havainnut lintujen ruoan ja nokan muodon yhteyden luonnossa." Esimerkki c-vastauksesta: "Ei, kokeiden tekemisen sijaan hän käytti muita menetelmiä asian tutkimiseen."

Tehtävä 1a testaa saman VASI-näkökulman (monet menetelmät) osaamista suoremmin ja siinä pyydetään perustellen vastaamaan kysymykseen "Pidätkö henkilön tekemää tutkimusta tieteellisenä?" Oppilasvastaukset olivat enimmäkseen n-vastauksia, esim. "Kyllä, koska siinä tutkittiin lintuja." Esimerkki m-vastauksesta: "Kyllä, koska hän on käyttänyt näköhavaintojaan apunaan." (Mainitaan yksi menetelmä.) Esimerkki c-vastauksesta: "Kyllä, koska hän on käyttänyt erilaisia menetelmiä asian tutkimiseen."

Tehtävä 1c testaa saman VASI-näkökulman (monet menetelmät) osaamista suorimmin: "Voidaanko luonnontieteellisessä tutkimuksessa mielestäsi käyttää useampaa kuin yhtä menetelmää?". Kieltävää vastausta pyydetään perustelemaan ja myöntävän vastauksen tapauksessa pyydetään kuvaamaan kaksi eri menetelmää. Tässä linjattiin niin, että kieltävä vastaus tai myöntävä vastaus ilman perusteluja on n-vastaus. Esimerkki m-vastauksesta: "Voidaan käyttää useampaa kuin yhtä. Voidaan ottaa kokeita ja sitten voidaan tutkia." Esimerkki c-vastauksesta: "Voidaan. Voidaan tutkia lintujen elintapoja ja ominaisuuksia tietyssä elinympäristössä ja sitten voidaan muuttaa niiden elinympäristöä ja tutkia miten ne pärjäävät." Toinen esimerkki c-vastauksesta: "Kyllä. Jos esim. tutkin erilaisten aineiden reagointia toistensa kanssa ja niiden reaktionopeuteen vaikuttavia tekijöitä, käytän erilaista menetelmää kuin tutkiessani eri muotoisten kappaleiden putoamiskiihtyvyyttä, vaikka kummassakin mitataan aikaa." (Nämä kaksi hyväksyttiin c-vastauksiksi, vaikkei näissäkään aivan tyhjentävästi kuvata käytettävien menetelmien eroa.)

2.5. Arvioinnin luotettavuus

Oppilaiden vastausten arviointia harjoiteltiin professori N. Ledermanin ryhmän kanssa (SKYPE-kokous). Vastausten koodaaminen pyrittiin muutenkin saamaan mahdollisimman objektiiviseksi käyttämällä kahta arvioijaa. Kahden arvioijan tulosten yhdenmukaisuuden ja samalla tulosten luotettavuuden arvioimiseksi laskettiin kahden arvioijan yksimielisyyskerroin, nk. Cohenin kappa (Cohen, 1960). Kerrointen laskennassa käytettiin neljänkymmenen satunnaisesti valitun oppilasvastauksen otosta. Taulukossa 4 esitetään kullekin VASI-näkökulmalle kyselyn tehtäväkohtien pisteytysten avulla laskettu keskimääräinen yksimielisyyskerroin K.

$$K = \frac{p_O - p_C}{1 - p_C} \qquad (1)$$

Jossa p_O = havaittu yksimielisyys ja p_C = todennäköisyys sille, että arvioijat päätyvät samaan arvioon sattumalta.

2.6. Otoskoko

Tutkimuksen aineisto kerättiin keskikokoisessa eteläsuomalaisessa peruskoulussa, jossa VASI-kyselyyn vastasi 149 seitsemäsluokkalaista. Projektin koordinaatioryhmän suositus otoskooksi oli vähintään 100 oppilasta. Etelä-Afrikassa lukiolaisilla tehdyn VASI-tutkimuksen otoskoko oli 105 (Gaigher, Lederman & Lederman, 2014).

Vuoden 2012 PISA-tuloksista nähtiin, että erityisesti matematiikan osaamisessa alueelliset ja koulukohtaiset erot ovat Suomessa vähäiset (Kupari ym. 2013). Tämän tuloksen johdosta voidaan varoen arvioida, että yhden isohkon koulun oppilaista muodostettu aineisto voi edustaa kohtuullisen hyvin suomalaisten seitsemäsluokkalaisten keskimääräistä osaamista.

3. Tulokset

VASI-tulokset esitetään koko aineistolle, luokittain (taulukko 5) ja kysymyskohtaisesti (taulukko 6) eriteltyinä ilmoittamalla mediaanit, tytöille ja pojille erikseen. Useimmilla luokilla tyttöjen tulos oli tilastollisesti merkitsevästi parempi kuin poikien, samoin kysymyskohtaisesti jokainen kysymys meni tytöillä paremmin kuin pojilla. Epäselviä ja puuttuvia vastauksia oli tyypillisesti muutamia. Edellä mainittu kahden arvioijan yksimielisyys oli kaikkien tehtävien kohdalla korkealla tasolla.

Taulukko 4. Yksimielisyyskerroin K kahdelle arvioijalle, kun N = 40.

VASI-näkökulma	K
A. Tutkimus alkaa tutkimuskysymyksellä	0,86
B. Ei ole yhtä oikeaa tutkimusmenetelmää	0,93
C. Kysymys ohjaa menetelmän valintaa	0,96
D. Sama menetelmä ei takaa samaa tulosta	0,91
E. Menetelmän valinta voi vaikuttaa tuloksiin	0,91
F. Johtopäätökset ovat linjassa aineiston kanssa	0,98
G. Johtopäätökset vs. aineisto	0,89
H. Johtopäätöksissä yhdistyvät aineisto ja tieto asioista	0,94

Voidaan ajatella, että mediaani 1 merkitsee naiivia osaamista ja mediaani 2 asiallisen osaamisen tyyppistä osaamista. Kokonaistuloksissa (taulukko 5) mediaani 2 esiintyy tytöillä viidellä luokalla ja pojilla yhdellä luokalla kahdeksasta. Mediaania 3 ei esiinny lainkaan taulukossa 5. Tämän tuloksen perusteella voidaan sanoa, että VASI-käsitteitä ei hallita seitsemännellä luokalla vielä kovin hyvin.

Taulukko 5. VASI-kokonaistulos luokittain, tytöt ja pojat erikseen

Luokka	$Md_{tytöt}$	n	Md_{pojat}	n
7a	2,00	6	2,00	9
7b	1,00	10	1,00	9
7c	1,00	12	1,00	8
7d	1,00	8	1,00	8
7e	2,00	8	1,00	14
7f	2,00	8	1,00	10
7g	2,00	10	1,00	14
7h	2,00	6	1,00	9
kaikki luokat	2,00	68	1,00	81

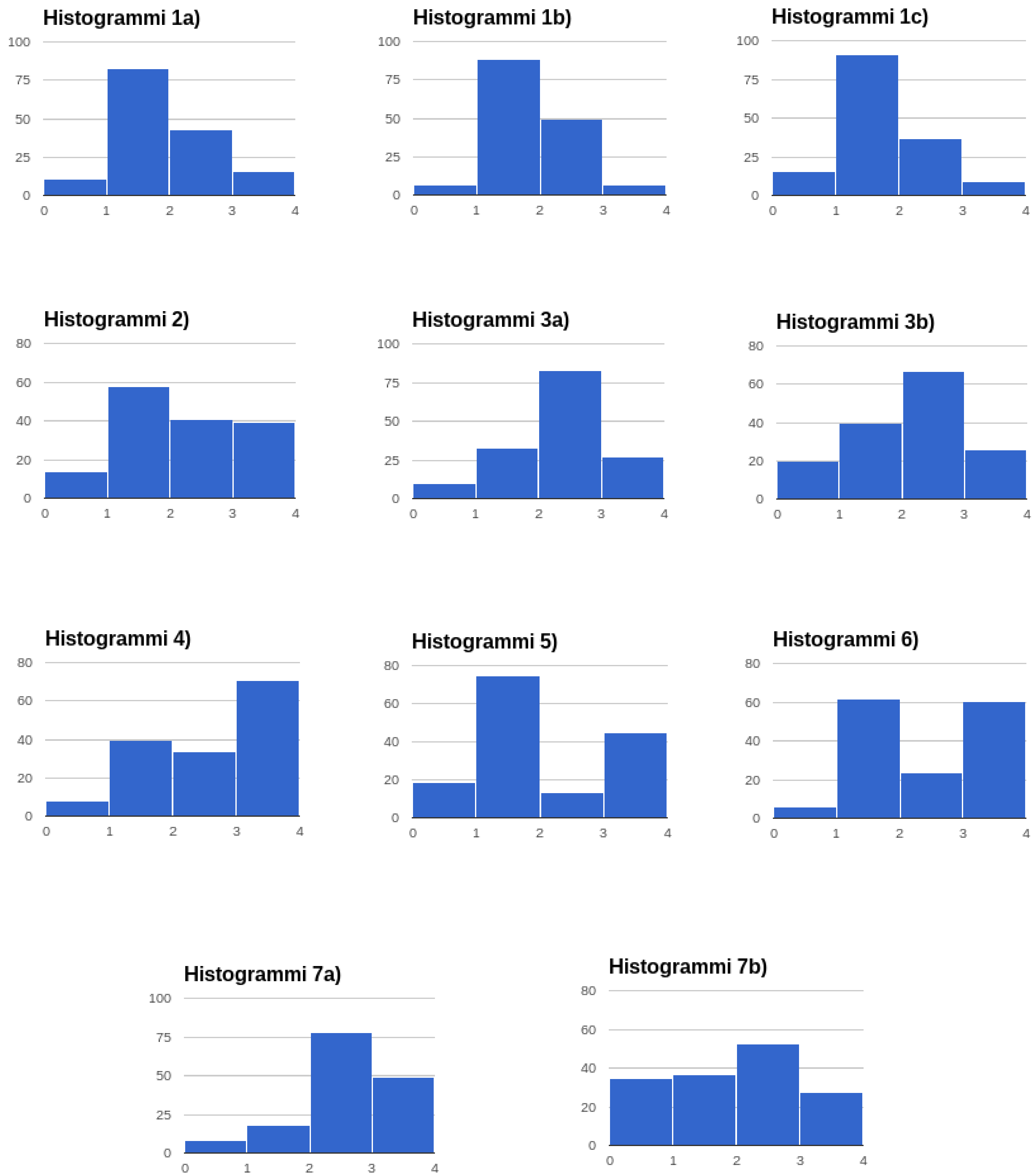
Tyttöjen ja poikien osaamisessa olevan eron tilastollista merkitsevyyttä testattiin Mann-Whitneyn U-testillä eli Wilcoxonin järjestyssummatestillä, joka on parametrisen t-testin epäparametrinen vastine (p- ja W-arvot on laskettu vapaasti ladattavalla R-ohjelmalla; merkitsevyys p lasketaan järjestyssummasta W). Tyttöjen osaaminen oli koko aineistossa ja osassa tehtävistä (tehtävät 4 ja 7a) tilastollisesti merkitsevästi ($p < 0,01$) poikien osaamista parempaa. Osassa kysymyksistä tyttöjen osaaminen oli melkein merkitsevästi ($p < 0,05$) poikien osaamista parempaa, kun taas osassa kysymyksistä poikien ja tyttöjen osaamisessa ei ollut tilastollisesti merkitsevää eroa.

Yleiskuva tehtäväkohtaisesta osaamisesta saadaan, kun tarkastellaan tyttöjen ja poikien tehtäväkohtaisia mediaaneja taulukossa 6. Tytöt ja pojat onnistuivat parhaiten tehtävissä 3, 4 ja 7a. Heikoimmin osattiin vastata tehtävissä 1 ja 5. Vähintään mediaaniin 2 tytöt ylsivät seitsemässä, pojat neljässä tehtävässä (yhdestätoista).

Taulukko 6. VASI-pisteiden mediaanit tehtäväkohtaisesti

VASI-tehtävä	VASI-näkökulma	$Md_{tytöt}$	Md_{pojat}	W	p
1a	B	1,00	1,00	2525	0,4
1b	B	1,00	1,00	2485	0,3
1c	B	1,00	1,00	2484	0,3
2	A	2,00	1,00	2245	0,06
3a	D	2,00	2,00	2318	0,09
3b	E	2,00	2,00	2272	0,07
4	G	3,00	2,00	2022	0,004
5	C	1,00	1,00	2458	0,3
6	F	2,00	1,00	2165	0,02
7a	H	2,00	2,00	2009	0,003
7b	H	2,00	1,00	2174	0,03

Mediaanien tarkastelu ei anna riittävän yksityiskohtaista kuvaa VASI-näkökulmien hallinnasta. Tehtäväkohtaisten pisteiden jakaumien tarkastelu antaa yksityiskohtaisempaa tietoa oppilaiden osaamisessa olevista eroista (kuva 2). Osaamiseen liittyvän tiedon kartoituksessa on mahdollista, jopa todennäköistä, että jakaumasta tulee kaksihuippuinen, jos näkemykset jakautuvat voimakkaasti tietyn c-vastauksen ja tietyn n-vastauksen välillä.



Kuva 2. VASI-kyselyn tehtäväkohtien pisteiden jakaumat histogrammeina

4. Pohdinta

Kysymykseen yksittäisten tehtävien hallinnasta saadaan vastaus yhdistämällä taulukon 6 ja kuvan 2 informaatiot. Huonosti osattujen tehtävien 1a-1c histogrammit ovat hyvin samanlaiset: vasemmalle vinojen jakaumien painopiste on ykkösen kohdalla. Sitä vastoin niin ikään huonosti osatun tehtävän 5 histogrammi on täysin erilainen (kaksihuippuinen): hyvin monet oppilaat olivat sitä mieltä, että on parempi tutkia yhdenlaisten renkaiden käyttäytymistä kolmella erilaisella tienpinnalla, vaikka tutkimuskysymys oli menevätkö tietyn tyyppiset renkaat helpommin rikki? A-ryhmän (virheellistä) tutkimusasetelmaa, jonka tavoitteena oli tutkia samanlaisten renkaiden käyttäytymistä kolmella erilaisella tienpinnalla, pidettiin mm. “laajempana” ja “monipuolisempana” kuin B-ryhmän, jolla pyrittiin saamaan vastaus suoraan esitettyyn tutkimuskysymykseen (ja vain siihen) eli erilaisten renkaiden käyttäytymiseen samanlaisella tienpinnalla.

Hiukan tehtävää 5 paremmin osatun tehtävän 6 histogrammi on myös kaksihuippuinen: kasvin saaman valon ja pituuskasvun yhteys kävi ilmi tehtävämonisteen taulukosta. Tästä huolimatta monet oppilaista valitsivat monivalintatehtävän kohdista virheellisen vaihtoehdon joko niin, että valitsivat tavanomaisen kokemusperäisen vaihtoehdon (kasvit tarvitsevat valoa), tai niin, että taulukon yksi muista poikkeava arvo esti päättelyn kokonaan. Siinä tapauksessa, että oppilasvastaukset polarisoituvat kahteen ääripäähän (c- ja n-luokat) kuten tehtävissä 5 ja 6, pelkkä mediaani ei kerro juuri mitään vaan tarvitaan arvosanojen jakauman kuvaus.

Tehtävien 5 ja 6 arvosanojen jakaumia verrattiin Etetä-Afrikassa lukiolaisilla tehdyn VASI-tutkimuksen tuloksiin (Gaigher, Lederman & Lederman, 2014). Etelä-afrikkalaisessa aineistossa tehtävät 5 ja 6 erottuivat selkeästi parhaiten menneinä tehtävinä: näissä kahdessa tehtävässä c-vastausten osuus oli n. 60%. Muissa tehtävissä keskimääräinen vastaus oli m-vastaus, joka vastaa mediaania 2 suomalaisessa aineistossa. Etelä-afrikkalaisessa aineistossa vain tehtävässä 5 jakaumassa oli kaksi huippua (n. 25% oli n-vastauksia). Suomalaisessa aineistossa tehtävissä 5 ja 6 havaittu jakaumien muoto voikin muuttua siirryttäessä yläkoulusta lukioon. Tällä havainnolla on ennustearvoa ja tätä voidaan testata myöhemmin, jos VASI-testi tehdään 9-luokkalaisilla tai lukiolaisilla. Tehtävien 5 ja 6 tapaisilla monivalintatehtävillä voidaankin ehkä mitata käsitteellistä oppimista tehokkaasti. Osaamisero suomalaisten yläkoululaisten ja etelä-afrikkalaisten lukiolaisten välillä oli suurimmillaan juuri tehtävissä 5 ja 6.

Verraten hyvin osatut tehtävät 3a ja 3b liittyvät keskenään läheisiin VASI-näkökulmiin (tutkimuskysymys ja -menetelmä vs. tulokset) ja testaavat näiden hallintaa melko suoraan, eikä ole yllätys, että tehtävien pisteiden oikealle vinot jakaumat ovat samanlaiset. Niiden mediaanit ovat kakkosen kohdalla. Tehtävät 4 ja 7a osattiin parhaiten: niiden pisteiden jakaumat olivat selkeästi oikealle vinot, mediaanit 2-3, näissä myös tyttöjen ja poikien osaamisen eron tilastollinen merkitsevyys oli suurimmillaan. Näissä kahdessa tehtävässä oppilaita rohkaistiin kertomaan omin sanoin ja käyttämään omaa luovuuttaan ja päättelykykyään. Tehtävään 4 liittyvä VASI-näkökulma (aineisto on eri asia kuin tulokset) osattiin ilman eksplisiittistä opetusta todennäköisesti siksi, että suomalaisilla oppilailla on keskimäärin hyvät valmiudet loogiseen päättelyyn ja riittävät kielelliset valmiudet arvioida käsitteiden välisiä eroja/yhtäläisyyksiä.

Tehtävän 2 pisteiden jakauma on vasemmalle vino. Väite luonnontieteellinen tutkimus alkaa aina tutkimuskysymyksellä osoittautui haastavaksi analysoida. Tässä olisi pitänyt ymmärtää, että tutkimustulosten saavuttaminen ilman tutkimuskysymystä on luonnontieteellisen tutkimuksen luonteen vastaista. On todennäköistä, että sana aina teki tästä tehtävästä erityisen vaikean.

Kysymystä siitä mitä osattiin ja miksi voidaan pohtia tietoteoreettisesti 2015 PISA-tutkimuksen luonnontieteiden osion tapaan: VASIssa nuoret osasivat analysoida ja selittää aineiston ja tulosten välisiä käsitteellisiä eroja (tehtävä 4), mutteivät osanneet valita kahdesta vaihtoehtoisesta menetelmästä oikeaa erilaisten autonrenkaiden tutkimiseksi (tehtävä 5). Tämän perusteella voitaisiin sanoa, että teoreettisempi käsite- ja sisältötieto oli paremmin hallussa kuin soveltavampi episteeminen ja menetelmätieto. Tyhjentävämpi tietoteoreettinen analyysi edellyttäisi laajempaa kysymysvalikoimaa. 2015 PISAn luonnontieteissä suomalaisnuoret osasivat selittää luonnontieteellisiä ilmiöitä hiukan paremmin kuin suunnitella ja arvioida tutkimuksia tai tulkita aineistoja ja tuloksia (Vettenranta ym., 2016).

Tyttöjen osaaminen oli koko aineistossa ja osassa tehtävistä tilastollisesti merkitsevästi poikien osaamista parempaa. Tämä tulos on yhteensopiva vuoden 2012 PISA-tulosten kanssa (Kupari ym., 2013). PISAssa suomalaisilla tytöillä pistemäärät lukutaidossa ja luonnontieteissä olivat keskimäärin 10 % korkeammat kuin pojilla, kun taas matematiikassa pisteet olivat lähes tasoissa. VASIssa mitattavan osaamisen voidaan ajatella jakautuvan sekä luonnontieteen että lukutaidon osaamisiin. Tulos on yhteensopiva myös vuoden 2015 PISA-tulosten kanssa (OECD, 2016).

5. Johtopäätökset

Johdannossa esitettiin neljä havaintoa aikaisemmista oppilaiden NOS-käsityksiä kartoittaneista tutkimuksista (Lederman, 2007). Kun käsillä olevia VASI-testin kokonaistuloksia (sekä tyttöjen että poikien osaaminen) tarkastellaan suhteessa näihin havaintoihin, ensimmäisen havainnon, ts. sen että kolmetoistavuotiaat eivät anna informoituja vastauksia, voidaan katsoa vahvistuneen. Muita kolmea havaintoa tämän diagnostisen VASI-testin tuloksista ei voida suoraan vahvistaa, koska NOS-opetusta ei annettu.

Tämän tuloksen perusteella luonnontieteiden luonnetta hahmottavalle eksplisiittiselle opetukselle olisi Khishfen ja Abd-El-Khalickin (2002) ehdottamalla tavalla tarve. Suurimmassa osassa kysymyksistä pisteiden jakaumat kertovat, että oppilailla ei ole asiallisia näkökulmia/käsityksiä vaan oppilaat vastaavat intuitiivisesti luonnontieteen luonteen osaamista kartoittaviin kysymyksiin. Se, että valtaosa oppilaista kuitenkin vastasi vaikeimpiinkin tehtäviin eikä jättänyt vastaamatta, voidaan hyväksyä signaalina kannustavasta ja lapsen positiivista minäkuvaa vahvistavasta opetuksesta.

Tutkimuksen tulokset ovat sangen ymmärrettäviä, kun niitä verrataan tutkimuksiin, joissa on selvitetty luonnontieteiden opetuksessa käytettäviä työtapoja. Esimerkiksi Lavonen ja Laaksonen (2009) vertailivat PISA-oppilaskyselyn perusteella oppilaiden kokemien työtapojen käytön määrää Suomessa muiden OECD-maiden oppilaiden kokemiin vastaaviin määriin. Suomalaiset oppilaat kokevat, että he tekevät perinteisiä ohjeen mukaan suoritettavia oppilastöitä enemmän kuin muiden OECD-maiden oppilaat. Sitä vastoin suomalaiset oppilaat asettavat tutkimuskysymyksiä, suunnittelevat tutkimuksia tai testaavat omia ideoitaan muiden OECD-maiden oppilaita huomattavasti vähemmän. Aineiston perusteella johtopäätöksiä tehdään Suomessa kuitenkin hivenen enemmän kuin muissa OECD-maissa. Samantapaisia havaintoja voidaan tehdä kansallisen ”Luonnontieteiden osaaminen perusopetuksen 9. luokalla 2011: Koulutuksen seurantaraportit 2012:2” perusteella (Kärnä, Hakonen, Kuusela, 2012).

Perusopetuksen luokilla 7 – 9 tulisikin osa perinteisistä oppilastöistä muokata sellaiseen muotoon, jossa oppilaita ohjataan asettamaan tutkimuskysymyksiä ja suunnittelemaan luonnontieteellisiä tutkimuksia. Tällaiset tilanteet ja niihin liittyvä keskustelu voisivat auttaa oppilasta hahmottamaan kokeellisen työskentelyn roolia osana luonnontieteellistä tutkimustyötä. Suunnittelussa on tärkeää tunnistaa, minkä muuttujien välistä yhteyttä halutaan selvittää ja mitä muuttujia vakioidaan ja mitä varioidaan. Tällaisen työskentelyn yhteydessä tulee nostaa esille keskustelussa tutkimuskysymyksen merkitys, tutkimusmenetelmän valinnan suhde asetettuun kysymykseen. Hyödyllistä on vertailla oppilaiden asettamia kysymyksiä, valitsemia tutkimusmenetelmiä ja saamia tuloksia. Kun opettaja ohjaa keskustelua, oppilaille avautuu tutkimuskysymyksen merkitys ja se kuinka tutkimuskysymys ohjaa menetelmän valintaa. Vertailemalla valittuja menetelmiä ja saatuja tuloksia opitaan, että ei ole yhtä oikeaa tutkimusmenetelmää ja että menetelmän valinta voi vaikuttaa tuloksiin ja toisaalta sama menetelmä ei takaa samaa tulosta. Vaikka Lavosen ja Laaksosen (2009) mukaan suomalaiset oppilaat kokevat tekevänsä muiden OECD-maiden oppilaita useammin aineistoon perustuvia johtopäätöksiä, olisi johtopäätösten laatuun kiinnitettävä huomiota. Johtopäätös edellyttää havaintoa tai mittausta ja aikaisempaa tietoa tai mallia, johon päätelmä kytketään. Myös tässä tarvitaan opettajan kyselyä ja ohjausta.

Kansainvälisessä VASI-tutkimuksessa kerätään käyttökokemuksia ja havaintoja eri maista. Tässä artikkelissa esitetään Suomen osalta saatuja tuloksia, pohdintaa ja havaintoja, joilla toivotaan olevan vaikutusta paitsi tutkimustyössä myös kyselyn hyödyntämisessä diagnostisissa ja/tai formatiivisissa tarkoituksissa suomalaisessa yläkoulussa.

Kiitokset

Kirjoittajat kiittävät lehtori Juha Lindqvistiä hänen osuudestaan oppilasvastausten arvioinnissa.

Lähteet

Abd-El-Khalick, F., Bell, R. & Lederman, N. (1998). The nature of science and instructional practice: Making the unnatural natural. Science Education, 82, 417–437.

Achieve, Inc. (2013). Next generation science standards. Retrieved June 25, 2013 from http://www.nextgenscience.org/next-generation-science-standards

Ahonen, S. (1994). Fenomenografinen tutkimus. Teoksessa L. Syrjälä, S. Ahonen, E. Syrjäläinen & S. Saari (toim.) Laadullisen tutkimuksen työtapoja. Helsinki: Kirjayhtymä Oy, 113–160.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.

Fairbrother, R. 2000. Strategies for learning. in M. Monk & J. Osborne (Eds.) Good practice in science teaching: What research has to say. Buckingham: Open University Press.

Gaigher, E., Lederman, N. & Lederman, J. (2014). Knowledge about inquiry: a study in South African high schools. International Journal of Science Education, 36, 3125-3147.

Khishfe, R. & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. Journal of Research in Science Teaching, 39, 551-578.

Kupari, P., Välijärvi, J., Andersson, L., Arffman, I., Nissinen, K., Puhakka, E. & Vettenranta, J. (2013). PISA 2012 ensituloksia. Opetus- ja kulttuuriministeriön julkaisuja 2013:20. Opetus- ja kulttuuriministeriö & Koulutuksen tutkimuslaitos. Jyväskylän yliopisto.

Kärnä, P., Hakonen, R. & Kuusela, J. (2012) Luonnontieteiden osaaminen perusopetuksen 9. luokalla 2011: Koulutuksen seurantaraportit 2012:2. Helsinki: Opetushallitus.

Lavonen, J. & Laaksonen, S. (2009). Context of teaching and learning school science in Finland: Reflections on PISA 2006 results. Journal of Research in Science Teaching, 46, 922-944.

Lederman, J., Lederman, N., Bartos, S., Bartels, S., Meyer, A. & Schwartz, R. (2014). Meaningful assessment of learners' understandings about scientific inquiry –

The views about scientific inquiry (VASI) questionnaire. Journal of Research in Science Teaching, 51, 65-83.

Lederman, N. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. Journal of Research in Science Teaching, 29, 331-359.

Lederman, N. (2007). Nature of science: Past, present, and future, in Handbook of research on science education, edited by S. Abell and N. Lederman, Lawrence Erlbaum Associates, Inc., Publishers, 2007 Mahwah, New Jersey.

Lederman, N., Lederman, J. & Antink, A. (2013). Nature of science and scientific inquiry as contexts for the learning of science and achievement of scientific literacy. International Journal of Education in Mathematics, Science and Technology, 1(3), 138-147.

Lederman, N. & Lederman, J. (2014). Reseach on teaching and learning of nature of science, in Handbook of research on science education, Volume II, edited by N. Lederman and S. Abell, Taylor and Francis Group Ltd, 2014, Oxford, UK.

National Research Council (2011). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Washington, D.C.: National Academy Press.

OECD, (2007). PISA 2006: Science competencies for tomorrow's world – Volume 1- Analysis. Paris, France: OECD Publishing.

OECD (2016). PISA 2015 Results (Vol. I): Excellence and Equity in Education, Paris, France: OECD Publishing. http://dx.doi.org/10.1787/9789264266490-en

Opetushallitus (1994). Perusopetuksen opetussuunnitelman perusteet 1994. Helsinki: Opetushallitus.

Opetushallitus (2004). Perusopetuksen opetussuunnitelman perusteet 2004. Helsinki: Opetushallitus.

Opetushallitus (2014). Perusopetuksen opetussuunnitelman perusteet 2014. Helsinki: Opetushallitus.

Palmu, T. (2003). Sukupuolen rakentuminen koulun kulttuurisissa teksteissä. Etnografia yläasteen äidinkielen oppitunneilla. Helsingin yliopiston kasvatustieteen laitoksen tutkimuksia 189. Helsinki: Yliopistopaino.

Schwartz, R., Lederman, N., & Lederman, J. (2008). An instrument to assess views of scientific inquiry: The VOSI Questionnaire. Paper presented at the International Conference of the National Association for Research in Science Teaching (NARST), Baltimore, MD.

Simon, S. 2000. Students attitudes towards science, in M. Monk & J. Osborne (Eds.) Good practice in science teaching: What research has to say. Buckingham: Open University Press.

Tala, S. & Vesterinen, V.-M. (2015). Nature of science contextualized: Studying nature of science with scientists. Science & Education 24, 435–457.

Vettenranta, J., Välijärvi, J., Ahonen, A., Hautamäki, J., Hiltunen, J., Leino, K., Lähteinen, S., Nissinen, K., Nissinen, V., Puhakka, E., Rautopuro, J. & Vainikainen, M-P. (2016). PISA 2015 ensituloksia. Opetus- ja kulttuuriministeriön julkaisuja 2016:41. Opetus- ja kulttuuriministeriö & Koulutuksen tutkimuslaitos. Jyväskylän yliopisto.

THE IMPACT OF TEACHER CHARACTERISTICS ON EDUCATIONAL DIFFERENTIATION PRACTICES IN LOWER SECONDARY MATHEMATICS INSTRUCTION

Ulrika Ekstam, Åbo Akademi, Faculty of Education and Welfare Studies
Karin Linnanmäki, Åbo Akademi, Faculty of Education and Welfare Studies
Pirjo Aunio, University of Helsinki, Department of Education

Abstract This study aimed to investigate how teachers' certification status, experience in instruction, and teachers' efficacy beliefs for teaching lower secondary students in mathematics are related to differentiated instruction practices. A total of 42 mathematics teachers and 27 special education teachers answered an electronic questionnaire regarding mathematics teaching efficacy beliefs and their frequency of use of differentiation practices. The results indicated that teachers' efficacy beliefs were related to differentiation in content, flexible examination models, and co-teaching. Neither certification status nor teacher experience in instruction was related to the frequency of use of differentiation practices. As teacher efficacy beliefs seem to have an effect on the use of differentiation practices, and especially on co-teaching, it should be important for teacher education to focus on developing pre-service teachers' efficacy beliefs as well as implementing a strong collaboration between different teacher groups.

Keywords: differentiation, educational support, mathematics, teacher efficacy beliefs

1 Introduction

Today's teachers are expected to be capable of adapting instruction to all kind of learners with different needs. This is also an important aspect of the Finnish National Core Curriculum, in which educational support highlights the possibilities of differentiation and differentiated instruction in the general classroom and is a strong component of all tiers of the Finnish three-tier educational support model (National Board of Education, 2011, 2015). Research on teacher quality has shown that there are several teacher characteristics, such as subject knowledge, certification status, experience in instruction and teacher efficacy beliefs, that affect instruction (Bolyard & Moyer-Packenham, 2008; van Garderen, Newman Thomas, Stormont, & Lembke, 2013; Hill, 2007; Maccini & Gagnon, 2006). Since teacher characteristics impact teaching strategies and instructions (Holzberger, Philipp, & Kunter, 2013; Midgley, Feldlaufer, & Eccles, 1989; Thoonen, Sleegers, Peetsma, & Oort, 2011), it can be assumed that teacher characteristics also impact the frequency of use of differentiated instruction in mathematics. To increase knowledge on how teacher characteristics affect teachers' use of differentiated instruction, this study will examine how the frequency of use of differentiation practices in mathematics instruction is related to teacher efficacy beliefs, certification status and experience in teaching mathematics to lower secondary students.

1.1 The three-tier educational support model

The Finnish three-tier support model for educational support focuses on early identification of students with learning difficulties and early intervention (National Board of Education, 2011, 2015). It has much in common with the *Responsiveness to Intervention* (RtI) model, widely used in the United States (Fuchs, Fuchs & Compton, 2012; Fuchs & Vaughn, 2012), although many differences exist according to both the theoretical and pedagogical frameworks (Björn, Aro, Koponen, Fuchs & Fuchs, 2016). RtI is defined as a school-wide process that integrates instruction, intervention and assessment, and that should be based on evidence-based research (Johnson & Smith, 2008) and has been developed to support low-performing students through early identification and multitier (commonly three-tier) intensified instruction (Lembke, Hampton & Beyers, 2012). The first tier provides instruction for all students and includes differentiated instruction and flexible student grouping as common instructional practices while the second tier offers additional educational support in additional small groups or in-class support for those students not responding to instruction in tier one. The third tier offers intensive instruction for students in need of more specialised support.

Assessment is an important part of RtI to guarantee students' gains and performances (Riccomini & Smith, 2011). However, as most of the research-based intervention programs focus on early grades, RtI has some challenges in secondary education (Johnson & Smith, 2008) caused by, for example, lack of school-wide processes and relevant assessment measures (Clarke, Lembke, Hampton, & Hendricker, 2011). The importance of teachers' professional skills for a successful implementation of RtI has also been noted (Brownell, Sindelar, Kiely & Danielson, 2010; Hoover & Patton, 2008).

In the Finnish three-tier support model the first tier, *general support*, aims to offer educational support as soon as possible to students not responding to average classroom education in order to prevent the rise of possible learning problems (Finnish National Board of Education, 2015). The target group for this level of support is the whole student population (Thuneberg et al., 2013a). General support can be organized by, for example, differentiated instruction, meaning that the teacher takes the students' diversity into account during instruction (Finnish National Board of Education, 2015). At this level of support shorter periods for tutoring outside the school day (i.e. remedial instruction) can be included and the support is most often provided by class or subject teachers.

If a student receiving general support continues to perform below expected levels, he or she is then provided *intensified support*, or the second tier of the three-tier support model (Finnish National Board of Education, 2015). Approximately 8% of the students in Finnish comprehensive education receive intensified support (Official Statistics of Finland, 2016). At this level of support, part-time special education in class (e.g. co-teaching) or outside the classroom (e.g. pull-out lessons) plays an important role (Björn, Aro & Koponen, 2015; Thuneberg et al., 2013b). The students' need of support should be evaluated regularly to be sure the support is efficient (Finnish National Board of Education, 2015).

Students who fail to respond to instruction during intensified support are evaluated (pedagogical statement), and an official decision concerning *special support* (tier three) is made by the school administrator (usually the principal) as necessary (Finnish National Board of Education, 2015; Thuneberg et al., 2013b). Students receiving special support should have an individual student plan and can study according to either a general or an individual curriculum in either general or special classes (Finnish National Board of Education, 2015).

In Finland, approximately 22% of all students receive (extra) educational support at some level (Official Statistics of Finland, 2016). In lower secondary education, mathematics is the subject in which the most students need extra support (Official Statistics of Finland, 2011). Furthermore, variations in students' mathematics performance tend to increase (Harju-Luukkainen & Nissinen, 2011; Metsämuuronen, 2013; Rautopuro, 2012). This means that mathematics teachers are more likely than other subject teachers to teach students who have been identified as low-performing and in need of extra educational support. Moreover, special education teachers working with lower secondary students deliver most of their instruction in mathematics (Takala, Pirttimaa, & Törmänen, 2009), meaning that, in addition to having knowledge in special education, they also need strong knowledge in how to support secondary students in mathematics.

1.2 Differentiated instruction

Differentiation can be defined as '... a systematic approach to planning curriculum and instruction for academically diverse learners' (Tomlinson & Strickland, 2005, p. 6). Teachers' ability to adjust educational work to respond to the varied needs of the student group is the basis for successful differentiated instruction (Kaonstantinou-Katzi et al., 2013). There are several elements of instruction that teachers can modify to support students in their learning process according to the students' readiness, interest and learning profile: content, process (i.e. methods of practice and performance), product (i.e. evaluation and assessment) and learning environment (Tomlinson & Strickland, 2005, UNESCO, 2004). Research has found that differentiated instruction has a positive impact on student learning and attitudes towards mathematics (Kaonstantinou-Katzi et al., 2013). Furthermore, students who receive instruction, are assigned tasks on a suitable personal level and, as a result, experience success are more likely to be motivated and to maintain their self-esteem (Tomlinson, 2008), which is important for students in need of support (Linnanmäki, 2002). Differentiated instruction requires teachers to have experience in different ways of teaching and learning, as well as strong knowledge of their students, including their backgrounds, experiences, interests and learning profiles (Kiley, 2011; Taylor, 2015; Tomlinson, 2008).

The National Council of Teachers of Mathematics (NCTM) (2000) has also recognized the need for differentiation, particularly in mathematics. Differentiated instruction is not a new idea; however, in the last decade, it has been more common among mathematics teachers. The teacher must be aware of what each individual student needs and plan instruction that takes different educational needs into account. To differentiate instruction in mathematics, a teacher may, for example, differ the learning tempo, the depth of the content and homework, the frequency of calculator and computer use, the time allotted for tests or to solve word problems, the use of different manipulative tools or flexible student groupings (National Board of Education, 2011, 2015; Tomlinson & Strickland, 2005). Differentiating mathematics teaching has been found to be more challenging in middle and high school than in earlier grades (Maccini & Gagnon, 2006; Mageira, Smith, Zigmond & Gebauer, 2005). This is mostly due to deep-rooted differences in students' mathematical levels, which begin as far back as kindergarten or grade 1 and continue to pose challenges for teachers throughout all grades (Small & Lin, 2010).

1.3 Teacher characteristics

Teacher characteristics are features that vary among teachers. They include, for example, teacher knowledge, certification, experience, attitudes and beliefs (e.g., Blömeke & Delaney, 2012; Clotfelter, Ladd & Vigdor, 2007; Holzberger et al., 2013; OECD, 2009). Teacher characteristics have typically been studied in the context of their relations with teacher effectiveness (Anthony & Walshaw, 2009; Holzberger et al., 2013; Kai, Caiser, Perry & Wong, 2009) and student achievement (e.g. Baumert et al., 2010). Traditionally, teacher characteristics have been measured by characteristics that are easy to measure and control such as certification, experience and subject knowledge. However, more recent research has shown that many other characteristics such as self-beliefs, motivation and interest have an impact on student performance, but they are much harder to measure (Hattie, 2015; Bong & Skaalvik, 2003; Bursal, 2010; Gresham, 2008; Kim, Sihn & Mitchell, 2014; Swars, 2005; Woodcock & Reupert, 2016).

Teacher certification

Earlier research has reported that teachers' certification status may have a positive effect on students' mathematics learning at all educational levels (Clotfelter et al., 2007; Neild, Farley Ripple, & Byrnes, 2009). However, the percentage of certified teachers, especially in mathematics, is noted to be lower in the middle grades than in high school (Neild et al., 2009; Kumpulainen, 2014). There are different requirements for certification, such as the level of educational degree, which vary between countries (see e.g. Ingersoll, 2007; Sahlberg, 2011; Wang, Coleman, Coley, & Phelps, 2003).

Earlier studies have found that teachers who major in mathematics or are certified to teach high school-level mathematics seems to have a greater positive correlation with students' mathematical achievement in middle school than teachers with primary or middle school certifications (or other complementary certifications; Clotfelter et al., 2007; Neild et al., 2009; Hill, 2007). Bouck (2005) also found that a low percentage of special education teachers in secondary education had proper pre-service training for instruction at this educational level, despite being certified for secondary education.

In Finland, to be certified to teach mathematics in lower secondary education, a teacher must have a master's degree with at least 60 ECTS (European Credit Transfer and Accumulation System) in mathematics and education (including teaching practice), as well as a major (120 ETCS) in another subject (if not mathematics). Furthermore, for a certification in special education (K–12), a teacher must have a master's degree with at least 60 ECTS in Special Education (including teaching practice) and a major in another subject (if not Special Education). It takes approximately five years to earn a teacher certification.

Teacher experience in instruction

In this study, teacher experience is defined as a teacher's cumulative experience in instruction. Several studies have reported that teacher experience has a positive impact on student achievement (Bolyard & Moyer-Packenham, 2008; Clotfelter et al., 2007; Harris & Sass, 2011); however, the results are somewhat inconsequential. Studies have found evidence of strong positive development at the beginning of teachers' careers; however, this tends to level off after 5 to 10 years (Bolyard & Moyer-Packenham, 2008, Feng & Sass, 2013; Harris & Sass, 2011). In addition, the positive effect of experience on student achievement is stronger for the middle and high school levels than for pre- and primary school (Bolyard & Moyer-Packenham, 2008). A study by Hill (2007) indicated that teachers with more experience in instruction performed better than novice teachers in mathematical teaching knowledge. Furthermore, middle school teachers who had experience teaching at the high school level reported having more mathematical teaching knowledge than teachers without such experience (Hill, 2007). Teachers' experience with diverse learners is also noted to have a positive effect on teachers' attitudes and beliefs (Subban & Sharma, 2005).

Self-efficacy

Self-efficacy has its origins in social cognitive theory, and it can be defined as a person's subjective perception of his or her capability to achieve a preferred outcome in a specific context (Bandura, 1977). Self-efficacy is formed through experiences and includes what individuals believe they can do with their existing skills, rather than the actual skills themselves (Bandura, 1977; Bong & Skaalvik, 2003). One's belief in one's own efficacy is developed through four main sources of influence: mastery experience, physiological factors, vicarious experiences and social persuasion (Bandura, 1994/1998). Of these factors, the most important contributing to an increase in self-efficacy is the experience of mastery: specifically, success increases self-efficacy, while failure decreases it. Bandura (1997) also found that self-efficacy is affected by processes and emotions that impact individuals' motivation and are skill-, task- and domain-specific. Furthermore, people with high beliefs in their capabilities usually approach difficult tasks as challenges to be mastered, rather than as threats to be avoided. Such an efficacious approach fosters deep interest and involvement in activities (Bandura, 1994/1998).

Teacher efficacy beliefs (i.e. teacher self-efficacy) are defined as teachers' beliefs and perceptions about their ability to teach students with varying needs and qualifications (Tschannen Moran, Woolfolk Hoy, & Hoy, 1998). They also include the teachers' beliefs about their ability to achieve desired student engagement and learning outcomes (Bandura, 1977, 1997; Skaalvik & Skaalvik, 2007). This means that teachers with high teacher efficacy beliefs trust his or her skills to instruct students with different needs, while teachers with low teacher efficacy beliefs are uncertain of his or her skills to teach students with varying needs. Earlier research has shown that teacher efficacy beliefs are connected to teachers' capability to organize and execute teaching tasks in specific contexts (Skaalvik & Skaalvik, 2007). Studies have also shown that teacher efficacy beliefs tend to vary between contexts and over time (Tschannen-Moran et al., 1998). Bandura (1997) reported that pre-service teachers and novice teachers establish their teacher efficacy beliefs at an early stage. He also indicated that, once established, teacher efficacy beliefs can be hard to change.

Teacher efficacy beliefs have been shown to be related to teaching strategies, instructions and motivation (Holzberger et al., 2013; Midgley, Feldlaufer, & Eccles, 1989; Thoonen et al., 2011) and to student achievement (Austin, 2013). Holzberger and colleagues (2013) reported that teachers with high efficacy beliefs tend to provide more student-centred instruction and invest more effort into implementing new teaching methods, strategies and personalised learning support. They also demonstrate greater flexibility in classroom engagement and lesson design (Temiz & Topeu, 2013). King-Sears and Baker (2014) indicated that teachers working with low-achieving students seem to benefit from having high self-beliefs, which help them maintain high levels of interest, motivation and beliefs in their own work.

Mathematics teaching efficacy can be defined as a teacher's belief in his or her ability to teach mathematics effectively (Enochs, Smith, & Huinker, 2000). Several studies have reported that mathematics teaching efficacy is a significant predictor of teachers' instructional strategies for mathematics and that teachers with high mathematics teaching efficacy are more effective in their teaching (Enochs et al., 2000; Gresham, 2008; Swars, 2005). Teachers' mathematics performance and mathematics self-efficacy have also been shown to be positively correlated with mathematics teaching efficacy (Bates, Latham & Kim, 2011; Newton, Evans, Leonard, & Eastburn, 2012; Swackhamer, Koeller, Basile, & Kimbrough, 2009). Furthermore, a teacher with high mathematics teaching efficacy is more likely to be deeply involved in student instruction and classroom engagement, as well as in the implementation of new teaching methods and strategies (Bates et al., 2011; Swackhamer et al., 2009; Takahashi, 2011; Temiz & Topeu, 2013).

1.4 Present study

Research on educational support in secondary mathematics is scarce. As a complement to the literature concerning factors affecting teacher differentiation in instruction, this study will focus on how teacher characteristics are related to the use of differentiation practices. The research questions are: How is teachers' use of differentiation practices in secondary education mathematics related to (1) teacher certification status, (2) teacher experience and (3) teacher efficacy beliefs.

2 Methods

2.1 Participants

The participants in this study were 42 mathematics teachers (21 women and 21 men) and 27 special education teachers (26 women and 1 man) working in Swedish-speaking lower secondary schools (students aged 13 to 15) in both rural and urban areas of Finland. Of the mathematics teachers, 71% were certified mathematics teachers and 74% had worked for five years or more. The mean age of the mathematics teachers was 43.3 years (age range: 25 to 63 years). Of the special education teachers, 72% were certified in special education, and 78% had worked for five years or more. The mean age of the special education teachers was 43.7 years (age range: 26 to 62 years).

2.2 Procedure and measure

This study was part of a research project targeting educational support for low-performing students in lower secondary mathematics education. An electronic questionnaire for special education teachers and mathematics teachers was sent to all of the principals of Swedish-speaking schools in Finland with education in grades 7 through 9 ($N = 55$). The principals were asked to forward the questionnaire to their schools' special education and mathematics teachers. Unfortunately, since it is impossible to know how many teachers actually received the questionnaire, response percentages could not be calculated.

The teachers' use of differentiated instruction was studied using a 5-point Likert-type scale (1 = *not at all* to 5 = *often*) for their frequency of use of nine differentiation practices: differentiation in content, use of calculators, manipulative tools, flexible examination models, part-time special education, homework support, complementary oral examinations, co-teaching and remedial education. The variables for the use of differentiation practices were one-item questions (see Table 2 in Appendix). The choice of differentiation practices for the questionnaire was based on suggestions for differentiation in educational support in the Finnish National Core Curriculum (National Board of Education, 2011, 2015).

There are several instruments for measuring teacher efficacy beliefs (e.g., The Teacher Sense of Efficacy Scale [Tschannen-Moran & Woolfolk Hoy, 2001] and Norwegian Teacher Self-Efficacy Scale [Skaalvik & Skaalvik, 2007]) but, since teacher efficacy is context- and situation-specific, and there is no such existing instrument in Swedish for the Finnish-Swedish population, a scale was constructed by author (and piloted on six teachers) for the particular purposes of this research project. The scale included eight items, for example 'I have enough knowledge about difficulties in mathematics and know what to do', addressing teacher efficacy beliefs, on which the teachers rated their perceived confidence in teaching low-performing students in mathematics. All items were answered on a 4-point Likert-type scale (1 = *strongly disagree* to 4 = *strongly agree*) and are shown in Table 2 (Appendix) (for more information about the scale, see Authors, 2017). The electronic questionnaire was sent to the schools in May 2013, and reminders were sent twice: once in June and once in August.

2.3 Data analysis

The data analyses were conducted in several stages. To measure teacher efficacy, earlier research with the used instrument has shown that a one-factor model (including five items, since a three of the items did not fit the model) of the teacher efficacy beliefs variable described the data best and gave an excellent model of fit ($x^2(5) = 5.45$, $p = .36$, CFI = 1.00, RMSEA = .04) (Authors, 2017). Furthermore, for the reliability confirmation Cronbach's alpha (five items) was calculated, and the result was acceptable (.82). The five items addressing teacher efficacy were recoded and summed to analyse of the teachers' efficacy beliefs levels. Correlations are shown in Table 2 (Appendix). A MANOVA was conducted to analyse the research questions and effect sizes, Cohen´s d, were calculated for the significant variables. The risk of committing a Type II error is elevated in studies with small sample size, thus, we decided to use a 90% level of significance to decrease the Type II error risk. This of course increases the risk of Type I error in the study but given the sample size, is arguably an acceptable trade-off.

The teachers were first divided into three groups based on their total teacher efficacy beliefs scores. The cutoff point was above 80th and lower than 20th percentile. Accordingly, the low teacher efficacy group had 13 teachers, the moderate teacher efficacy group had 39 teachers, and high teacher efficacy group had 13 teachers, four teachers did not answer all the items for teacher efficacy beliefs. Of the 14 *non-certified* teachers, six (43%) were in the low teacher efficacy group, seven (50%) were in the moderate group and one (7%) was in the high group of teacher efficacy beliefs. Of the 51 *certified* teachers, 13% (7) were included in the low level group, 63% (32) were included in the moderate level group and 24% (12) were in the high level group (Figure 1). The distribution for *teacher experience* (< 10y and ≥10 y) was similar for teacher experience: approximately 20% (low), 60% (moderate) and 20% (high), respectively.

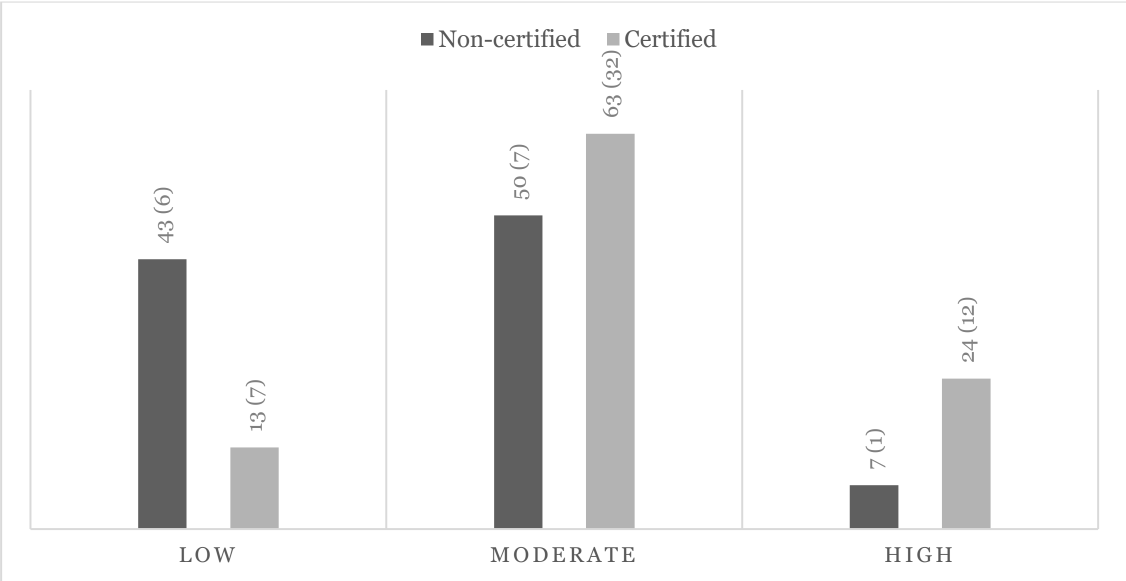


Figure 1. *Distribution (% and frequency) of certification status across different levels of teacher efficacy beliefs.*

3 Results

To analyse how teacher efficacy beliefs, certification status and teaching experience were related to the frequency of use of differentiation practices in mathematics, a MANOVA test was conducted. First, the three variables of teacher efficacy beliefs, experience and certification were tested separately. The preliminary results (Pillai's Trace) showed no significant (p >.1) differences for years of experience and certification status (p = .990 and p = .901) on differentiation practices. However, the p-value for teacher efficacy beliefs was .100 (η^2 = .214), and the between-subject test showed significant differences between teacher efficacy beliefs groups for three of the differentiation practices: *differentiation in content* (F(2,61) = 5.681, p = .006, η^2 = .166), *flexible examination models* (F(2, 61) = 2.461, p = .094,

η^2 = .079) and *co-teaching* (F(2, 61) = 4.543, p = .015, η^2 = .137). The results for all other variables were non-significant: use of calculators (p = .334), part-time special education (p = .490), homework support (p = .156), complementary oral examinations (p = .431) and remedial education (p = .848). Means for the three groups of teacher efficacy beliefs for the nine differentiation practices are shown in Figure 2.

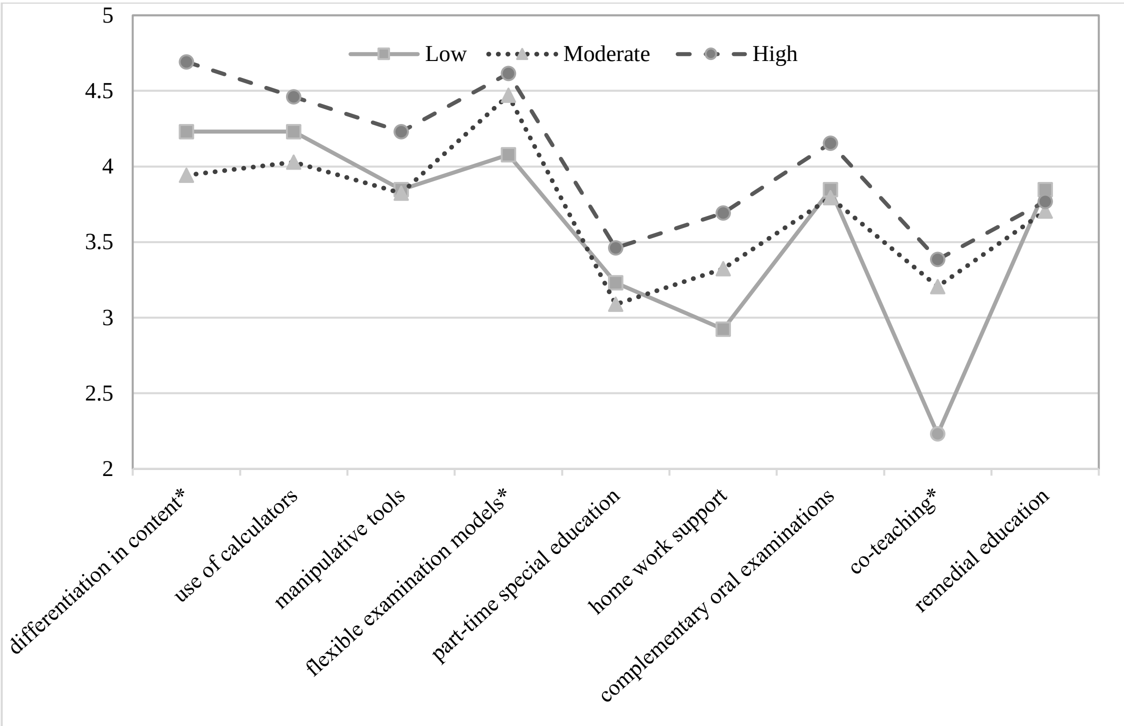


Figure 2. *Frequency of use (1 = not at all* to *5 = often) of differentiation practices for levels of teacher efficacy beliefs (low, moderate and high). Significant differences (*) were noted for differentiation in content, manipulative tools, and co-teaching*

To examine how the different groups of teacher efficacy beliefs differed for the significant variables, a post-hoc test was conducted. For differentiation in content there was a significant difference between the high and the other two (moderate and low) groups of teacher efficacy beliefs ($p_{\text{low-high}}$ =.093, d = .748 and $p_{\text{moderate-high}}$ = .001, d = 1.207). The results for flexible examination models showed that the high and moderate groups of teacher efficacy had a significant more frequent use than the teachers in the low teacher efficacy group ($p_{\text{low-high}}$ =.041, d = .835 and $p_{\text{low-moderate}}$ = .071, d = .553). The results also indicated that teachers with low teacher efficacy beliefs used co-teaching significantly less than teachers with high ($p_{\text{low-high}}$ = .010, d = .978) and moderate ($p_{\text{low-moderate}}$ = .009, d = .951) teacher efficacy beliefs. Means and standard deviations for all differentiation practices are shown in Table 1 (Appendix).

4 Discussion

This study examined how teacher efficacy beliefs (low, moderate and high), certification status (certified or non-certified) and teaching experience (years of teaching) are related to the frequency of use of educational differentiation practices in lower secondary mathematics instruction. The results indicated that teachers with high teacher efficacy beliefs for teaching mathematics to low-performing students are statistically significantly more likely to use differentiation in content, flexible examination models, and co-teaching than teachers with low teacher efficacy beliefs. With respect to teacher experience and certification, no notable significant differences between years of experience and certification status on groups of teacher efficacy beliefs were found. However, nearly half of the non-certified teachers belonged to the group with low teacher efficacy beliefs, while only 13% of the certified teachers belonged to this group. About 20% of the teachers had high teacher efficacy beliefs, approximately 60% had moderate and approximately 20% had low teacher efficacy beliefs, no matter of years of experience.

Earlier research has reported that teacher characteristics are related to, for example, teaching practices and student achievement (Anthony & Walshaw, 2009; Austin, 2013; Holzberger et al., 2013; Kai, Caiser, Perry & Wong, 2009; OECD, 2009). In this study, teachers with high teacher's efficacy beliefs was found to more frequency use differentiation in content, flexible examination models, and co-teaching. These findings are in line with earlier research which indicated that teachers with high teacher efficacy beliefs invest more effort into implementing new teaching methods, strategies and personalised learning support (Holzberger et al., 2013), while also demonstrating greater flexibility in classroom engagement and lesson design (Temiz & Topeu, 2013), all of which are important for effective differentiation (Tomlinson & Strickland, 2005). Furthermore, it has been reported that teachers with high teacher efficacy beliefs are more capable of organizing and executing teaching tasks for specific contexts (Skaalvik & Skaalvik, 2007).

Teachers with high teacher efficacy beliefs were significantly more frequently using differentiation in content and flexible examination models than teachers with moderate and low teachers' efficacy beliefs, while teachers with low teacher efficacy beliefs reported significant lower use of co-teaching than both teachers with moderate or high teacher efficacy beliefs. Since co-teaching is considered to be an effective model for support in secondary mathematics (Friend, 2008; Mageira et al., 2005) this result is pertinent to the development of teacher efficacy beliefs. Earlier research has also found that teacher efficacy beliefs is a significant predictor of teachers' instructional strategies for mathematics (Bates et al., 2011; Swackhamer et al., 2009; Takahashi, 2011; Temiz & Topeu, 2013), and since teachers with high mathematics teaching efficacy are found to be more effective in their teaching (Enochs et al., 2000; Gresham, 2008; Swars, 2005), it seems clear that teacher efficacy beliefs are important for achieving successful differentiated instruction in mathematics. Since teaching efficacy is reported to be context and subject specific and developed in an early stage (Bandura, 1997) the foundation for high mathematics teacher efficacy beliefs must start already during teacher education.

A highly qualified teacher has full certifications and demonstrates competence in both subject knowledge and teaching skills, use a wide range of learning strategies and understand how students can learn mathematics (U.S. Department of Education, 2002). Teacher certification has earlier been noted to impact both teaching strategies and instruction practices in mathematics (Feng & Sass, 2013; National Research Council, 2000). In this study, certification was not related to any of the differentiation practices. However, it is worth noting that almost half the non-certified teachers belonged to the group of low teacher efficacy beliefs, which was found to be related to differentiated instruction, and that the overall number of non-certified teachers was low.

In this study, a majority of the teachers had worked more than five years. The results from present study indicated that teacher experience was not statistically significantly related to the use of differentiation practices. Foss and Kleinsasser (1997) indicated that due to a lack of teaching experience, inexperienced teachers (e.g., pre-service teachers) tend to overestimate their teacher efficacy beliefs for teaching low-performing students, which can partly be reason for the non-significant results.

For six of the nine differentiation practices (use of calculators, use of manipulative tools, part-time special education, complementary oral examinations, homework support, and remedial education), no significant relation was found between groups of teacher efficacy and the use of differentiation practices. One reason for this lack of a significant relation may be the high number of students provided educational support in Finnish schools (Official Statistics of Finland, 2016). According to Official Statistics of Finland (2016), in 2015, nearly 23% of the students in compulsory education were provided general, intensified or special support, and approximately 80% (in Swedish-speaking schools) of the students receiving special support (full-time special education) were, at some point, included in the general classroom. As a result, it is likely that Finnish teachers, especially in Swedish speaking schools, have relatively high levels of experience with instructional differentiation and are able to effectively use the most traditional differentiation practices common in lower secondary mathematics. It is also worth noting that several of those differentiation practices with no statistically significant relation (e.g., part-time special education, complementary oral examinations, homework support, and remedial education) are related to the differentiation of product (i.e. evaluation and assessment) and learning environment, which may be are easily implemented by all teachers. Most of the differentiation practices analysed in this study with significant differences for level of teacher efficacy beliefs (differentiation in content, flexible examination models and co-teaching) all require both high subject knowledge in mathematics and confidence and interest in teaching low-performing students, both of which have been shown to impact teacher efficacy beliefs (Clotfelter et al., 2007; Holzberger et al., 2013; Kleinsasser, 2014).

4.1 Limitations

This study has some limitations. First, the number of participants is low. With more participants, the statistical results would have been stronger, and some of the differences that were close to significant (e.g., certification status, homework support, and manipulative tools) may have been significant. The study validation could also have been improved using a standardized scale for teacher efficacy beliefs. Furthermore, it would have been useful to include more items on differentiated instruction practices and *how* teachers differentiate in practice (e.g. content). For self-reported data, there are two issues that must be considered according to the validity. First, the cognitive factor (weather the respondents understand the questions and weather they have the knowledge to answer it) and second, the situational factor (the influence of the setting of the survey) (Brener, Bill & Grady, 2003). In this study both the cognitive (this study was part of a larger project concerning educational support in mathematics, so the meaning of the questions for the respondents was clear) and the situational (the respondents could answer anonymously when they had time). However, it is important to take into account that answering this questionnaire was voluntary, and therefore it is difficult to get a sample of participants representing a whole teacher group (Wright, 2005)

4.2 Conclusions

Instructional differentiation is an important part of educational support in secondary mathematics, especially for low-performing students in general education (Finnish National Board of Education, 2011, 2015; NCTM, 2000). This study examined how teacher efficacy beliefs, teacher experience and teacher certification affect the use of differentiation practices in mathematics instruction. The results indicated that teacher efficacy beliefs are important for several differentiation practices, especially those that focus on content and process, for example co-teaching which was found to be used significantly more frequently by both moderate and high teacher efficacy groups than by the low teacher efficacy group. Since teacher efficacy beliefs are established at an early stage of teacher education (Bandura, 1997), the foundation for teacher efficacy beliefs must also begin during teacher education. Thus, it is important to examine how teacher education can support pre-service teachers in developing high teacher efficacy beliefs for teaching mathematics to low-performing students. This concerns both special and mathematics pre-service teachers. Teacher education should also focus on how to support in-service teachers in both special education and mathematics to strengthen their teacher efficacy beliefs for teaching low-performing students in mathematics. This could be realized by introducing and implementing collaboration between different pre-teacher groups (e.g., mathematics and special education pre-teachers), learning how to use teachers specialized (subject and pedagogical) knowledge in a fruitful way for developing the most efficient educational support models and practices.

References

Anthony, G. & Walshaw, M. (2009). Characteristics of effective teaching of mathematics: A view from the West. *Journal of Mathematics Education*, *2*(2), 147–164.

Austin, J. (2013). The role of contextual factors in understanding mathematics teacher efficacy beliefs. *Journal of Mathematical Science & Mathematics Education*, *8*(2), 46–60.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215.

Bandura, A. (1994/1998). Self-efficacy. In H. Friedman (Ed.), *Encyclopedia of mental health*. San Diego: Academic Press. (Reprinted from *Encyclopedia of human health*, Vol. 4, pp. 71–81, by V. S. Ramachaudran, Ed., 1994, New York: Academic Press). Retrieved from https://www.uky.edu/~eushe2/Bandura/BanEncy.html.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bates, A., Latham, N., & Kim, J. (2011). Linking preservice teachers' mathematics self-efficacy and mathematics teaching efficacy to their mathematical performance. *School Science and Mathematics*, *111*(7), 325–333.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180. doi:10.3102/0002831209345157

Björn, P. M., Aro, & Koponen, M. (2015). Interventiovastemallien tarjoamat mahdollisuudet kolmiportaisen tuen kehittämiseen: esimerkkinä matematiikan oppimisen tuki [Developing interventionmodels for the three-tier support in mathematics]. *NMI-Bulletin*, *25*(3), 10-21.

Björn, P. M., Aro, M., Koponen, T., Fuchs, L. S., & Fuchs, D. H. (2016). The Many Faces of Special Education Within RTI Frameworks in the United States and Finland. *Learning Disability Quarterly*, 39 (1), 58-66. doi:10.1177/0731948715594787

Blömeke, S. & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM Mathematics Education*, *44*, 223–247.

Bolyard, J., & Moyer Packenham, P. (2008). A review of the literature on mathematics and science teacher quality. *Peabody Journal of Education*, *83*(4), 509–535.

Bong, M., & Skaalvik, E. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1–40. doi:10.1023/A:1021302408382

Bouck, E. (2005). Secondary special educators: Perspectives of preservice preparation and satisfaction. *Teacher Education and Special Education, 28*(2), 125–140.

Brener, N., Billy, J. O. G., & Grady, W. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health, 33*(6), 436-457. doi:10.1016/S1054-139X(03)00052-1

Brownell, M., Sindelar, P., Kiely, M., & Danielson, L. (2010). Special education teacher quality and preparation: Exposing foundations, constructing a new model. *Exceptional Children, 76*(3), 357–377.

Bursal, M. (2010). Turkish preservice elementary teachers' self-efficacy beliefs regarding mathematics and science teaching. *International Journal of Science and Mathematics Education*, *8*(4), 649–666.

Cai, J., Kaiser, G., Perry, B., & Wong, N. Y. (Eds.), *Effective mathematics teaching from teachers' perspectives: National and crossnational studies*. Rotterdam, The Netherlands: Sense Publishers.

Clarke, B., E. S. Lembke, D. D. Hampton, & Hendricker, E. (2011). What We Need to Know About Measuring Student Response in Mathematics. In Gersten, R., & Newman-Gonchar, R. (Eds.) *Understanding RTI in Mathematics. Proven Methods and Applications* (pp. 35–48). Baltimore: Brookes Publishing Company.

Clotfelter, C., Ladd, H., & Vigdor, J. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682. doi:10.1016/j.econedurev.2007.10.002

Enochs, L., Smith, P., & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and Mathematics*, *100*(4), 194–202. doi:10.1111/j.1949-8594.2000.tb17256.x

Feng, L., & Sass, T. (2013). What makes special-education teachers special? Teacher training and achievement of students with disabilities. *Economics of Education Review*, *36*, 122–134.

Finnish National Board of Education. (2011). *Ändringar och kompletteringar av grunderna för läroplanen för den grundläggande utbildningen 2010* [Changes and complements in national core curriculum for basic education]. Vammala, Finland: National Board of Education.

Finnish National Board of Education. (2015). *Grunderna för läroplanen för den grundläggande utbildningen 2014* [National core curriculum for basic education 2014]. Tampere: National Board of Education.

Foss, D., & Kleinsasser, R. (1996). Preservice elementary teachers' views of pedagogical and mathematical content knowledge. *Teaching and Teacher Education*, *12*(4), 429–442.

Friend, M. (2008). Co-teaching: A Simple Solution That Isn't Simple After All. *Journal of Curriculum and Instruction, 2,* 9-19.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). Smart RTI: A Next Generation Approach to Multilevel Prevention. *Exceptional Children* 78 (3): 263–279.

Fuchs, L.S., & Vaughn, S. (2012). Responsiveness-to-Intervention: A Decade Later. *Journal of Learning Disabilities* 45 (3): 195–203. doi: 10.1177/0022219412442150.

Gresham, G. (2008). Mathematics anxiety and mathematics teacher efficacy in elementary pre-service teachers. *Teaching Education*, *19*(3), 171–184. doi:10.1080/10476210802250133

Harju-Luukkainen, H., & Nissinen, K. (2011). *Finlandssvenska 15-åriga elevers resultatnivå i PISA 2009—undersökning* [Swedish speaking 15-year-old students' skills in mathematics in Finland in PISA 2009]. Jyväskylä: Koulutuksen tutkimuslaitos, University of Jyväskylä.

Harris, D., & Sass, T. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics, 95*(7), 798–812. doi:10.1016/j.jpubeco.2010.11.009

Hattie, J. (2015). The applicability of visible learning to higher education. *Scholarship of Teaching and Learning in Psychology, 1*(1), 79-91. doi:10.1037/stl0000021

Hill, H. (2007). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind Policy Initiative. *Educational Evaluation and Policy Analysis*, *29*(2), 95–114. doi:10.3102/0162373707301711

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology*, *105*(3), 774–786.

Hoover, J. J., & Patton, J. R. (2008). The role of special educators in a multitiered instructional system. *Intervention in School and Clinic*, *43*(4), 195–202. http://unesdoc.unesco.org/images/0014/001402/140224e.pdf

Ingersoll, R. M. (Ed.). (2007). *A comparative study of teacher preparation and qualifications in six nations*. Consortium for Policy Research in Education. Retrieved from http://www.cpre.org/images/stories/cpre_pdfs/sixnations_final.pdf

Johnson, E., & Smith, L. (2008). Implementation of response to intervention at middle school: Challenges and potential benefits. *Teaching Exceptional Children, 40*(3), 46–52.

Kiley, D. (2011). *Differentiated Instruction in the Secondary Classroom: Analysis of the Level of Implementation and Factors that Influence Practice*. *Dissertations*. 427. http://scholarworks.wmich.edu/dissertations/427

Kim, R., Sihn, H., & Mitchell, R. (2014). South Korean elementary teachers' mathematics teaching efficacy beliefs implications for educational policy and research. *Mathematics Education Trends and Research*, *2014*, 1–17. doi:10.5899/2014/metr-00052

King-Sears, M. & Baker, P. H. (2014). Comparison of teacher motivation for mathematics and special educators in middle schools that have and have not achieved AYP. *ISRN Education, Article ID 790179*, 1–12. doi:10.1155/2014/790179

Kleinsasser, R. (2014). Teacher efficacy in teaching and teacher education. *Teaching and Teacher Education*, *44* 168–179.

Konstantinou-Katzi, P., E. Tsolaki, M. Meletiou-Mavrotheris, & M. Koutselini. (2013). Differentiation of teaching and learning mathematics: An action research study in tertiary education. *International Journal of Mathematical Education and Technology*, *44*(3): 332–349. doi:10.1080/0020739X.2012.714491.

Kumpulainen, T. (Ed.). (2014). *Opettajat Suomessa* [Teachers in Finland]. *Koulutuksen seurantaraportit 2014: 8*. Tampere, Finland: National Board of Education.

Lembke, E., Hampton, D., & Beyers, S. (2012). Response to intervention in mathematics: Critical elements. *Psychology in the Schools, 49*(3), 257–272.

Linnanmäki, K. (2002). *Matematikprestationer och självuppfattning* [Mathematical achievement and self-concept]. Åbo: Åbo Akademis förlag.

Maccini, P., & Gagnon, J. (2006). Mathematics instructional practices and assessment accommodations by secondary special and general educators. *Exceptional Children, 72*(2), 217–234.

Mageira, K., C. Smith, N. Zigmund, & Gebauer, K. (2005). Benefits of co-teaching in secondary mathematics classes. *Teaching Exceptional Children*, *37*(3), 20–24.

Metsämuuronen J. 2013. Matemaattisen osaamisen muutos perusopetuksen luokilla 3–9 [The change in mathematical skills in grades 3–9]. In J. Metsämuuronen (Ed.), *Perusopetuksen matematiikan oppimistulosten pitkittäisarviointi vuosina 2005–2012* [Longitudinal Research of Student Mathematics Performance in Finnish Compulsory Education *2005–2012*] (pp. 65–172). Koulutuksen seurantaraportit 2013: 4. National Board of Education. Tampere: Juvenes Print—Suomen Yliopistopaino Oy.

Midgley, C., Feldlaufer, H., & Eccles, J. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during transition to junior high school. *Journal of Educational Psychology*, *81*(2), 247–258.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

National Research Council. (2002). *Educating teachers of science, mathematics and technology. New practices for the new millennium*. Retrieved from https://www.nap.edu/read/9832/chapter/5#45

Neild, R., Farley Ripple, E., & Byrnes, V. (2009). The effect of teacher certification on middle grades achievement in an urban district. *Educational Policy*, *23*(5), 732–760.

Newton, K., Leonard, J., Evans, B., & Eastburn, J. (2012). Preservice elementary teachers' mathematics content knowledge and teacher efficacy. *School Science and Mathematics*, *112*(5), 289–299.

OECD. (2009). *Teachers' pedagogical knowledge and the teaching profession*. Retrieved from http://www.oecd.org/edu/ceri/Background_document_to_Symposium_ITEL-FI-NAL.pdf

Official Statistics of Finland. (2011). *Special education. Appendix table 5. Comprehensive school pupils receiving part-time special education in the academic year 2009–2010 by primary reason for special education*. Retrieved from http://www.stat.fi/til/erop/2010/erop_2010_2011-06-09_tau_005_en.html

Official Statistics of Finland. (2016). *Special education*. Helsinki: Statistics Finland. Retrieved from http://www.stat.fi/til/erop/index_en.html

Rautopuro J. (Ed.). (2013). *Hyödyllinen pakkolasku. Matematiikan oppimistulokset peruskoulun päättövaiheessa 2012* [Mathematical skills at the end of comprehensive school 2012]. Koulutuksen seurantaraportit 2013:3. Opetushallitus. Helsinki: Juvenes Print—Suomen Yliopistopaino.

Riccomini, J. P., & Smith, G. W. (2011). Introduction of Response to Intervention in Mathematics. In *Understanding RTI in Mathematics. Proven Methods and Applications*, edited by In R. Gersten & R. Newman-Gonchar, (pp. 1–16). Baltimore: Brookes Publishing Company.

Skaalvik, E., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, *99*(3), 611–625.

Sahlberg, P. (2011). Lessons from Finland. *The Education Digest, 77*(3), 18–24.

Small, M., & Lin, A. (2010). *More good questions: Great ways to differentiate secondary mathematics instruction*. New York and Reston, VA: Teachers College Press and the National Council of Teachers of Mathematics.

Subban, P., & Sharma, U. (2005). Understanding educator attitudes toward the implementation of inclusive education. *Disability Studies Quarterly, 25*(2). Retrieved from http://dsq-sds.org/article/view/545/722

Swackhamer, L., Koellner, K., Basile, C., & Kimbrough, D. (2009). Increasing the self-efficacy of inservice teachers through content knowledge. *Teacher Education Quarterly, 36*(2), 63–78.

Swars, S. (2005). Examining perceptions of mathematics teaching effectiveness among elementary preservice teachers with differing levels of mathematics teacher efficacy. *Journal of Instructional Psychology*, *32*(2), 139–147.

Takahashi, S. (2011). Co-constructing efficacy: A 'communities of practice' perspective on teachers' efficacy beliefs. *Teaching and Teacher Education*, *27*(4), 732–741.

Takala, M., R. Pirttimaa, & M. Törmänen. (2009). Inclusion in special education—The role of special education teachers in Finland. *British Journal of Special Education*, *36*(3): 162–173.

Taylor, B. K. (2015). Content, Process, and Product: Modeling Differentiated Instruction. *Kappa Delta Pi Record*, *51*(1), 13-17. doi: 10.1080/00228958.2015.988559

Temiz, T., & Topeu, M. (2013). Preservice teachers' teacher efficacy beliefs and constructivist-based teaching practice. *European Journal of Psychology of Education*, *28*(4), 1435–1452.

Thoonen, E. E. J., Sleegers, P. J. C., Peetsma, T. T. D., & Oort, F. (2011). Can teachers motivate students to learn? *Educational Studies*, *37*(3), 345–360.

Thuneberg, H., Hautamäki, J., Ahtiainen, R., Lintuvuori, M.,Vainikainen M-P., & Hilasvuori, T.O. (2013a). Conceptual Change in Adopting the Nationwide Special Education Strategy in Finland. *Journal of Educational Change 15*: 37-56. doi: 10.1007/s10833-013-9213-x.

Thuneberg, H., Vainikainen, M-P., Ahtiainen, R., Lintuvuori, M., Salo, K., & Hautamäki, J. (2013b.) Education is Special for All – The Finnish Support Model. *Gemeinsam Leben* 21 (2): 67–78.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing and elusive construct. *Teaching and Teacher Education, 17,* 783-805.

Tschannen Moran, M., Woolfolk Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*(2), 202–248.

Tomlinson, C. A. (2008). The goals of differentiation. *Educational Leadership*, *66*(3): 26–30.

Tomlinson, C. A., & Strickland, C. A. (2008). *Differentiation in practice: A resource guide for differentiating curriculum, grades 9-12*. Alexandria, VA: Association for Supervision and Curriculum Development.

UNESCO. (2004). *Changing teaching practices, using curriculum differentiation to respond to students' diversity.* Paris: UNESCO. Retrieved from http://unesdoc.unesco.org/images/0013/001365/136583e.pdf

U.S. Department of Education. (2002). *No Child Left Behind Act 2001*. Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/index.html

van Garderen, D., Thomas, C., Stormont, M., & Lembke, E. (2013). An overview of principles for special educators to guide mathematics instruction. *Intervention in School and Clinic*, *48*(3), 131–141. doi:10.1177/1053451212454006

Wang, A., Coleman, A. B., Coley, R. J., & Phelps, R. (2003). *Preparing teachers around the world*. Princeton, NJ: Educational Testing Service.

Woodcock, S., & Reupert, A.E. (2016). Inclusion, classroom management and teacher efficacy in an Australian context. In S. Garvis, & D. Pendergast (Eds.), *Asia-Pacific Perspectives on Teacher Self-Efficacy*. (pp. 87 - 102). Rotterdam Netherlands: Sense Publishers.

Wright, K. (2005). Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*(3). doi: 10.1111/j.1083-6101.2005.tb00259.x

APPENDIX. Questionnaires.

Table 1. Items, mean and standard deviation for the different groups´ use of differentiation practices (1 = not at all to 5 = often); low, moderate and high teacher efficacy beliefs.

	M (SD)		
	Low	**Moderate**	**High**
Differentiation in content	4.23 (0.72)	3.94 (0.74)	4.69 (0.48)
Use of calculators	4.23 (0.73)	4.02 (1.03)	4.46 (0.66)
Manipulative tools	3.84 (0.55)	3.82 (0.83)	4.23(0.93)
Flexible examination models	4.08 (0.76)	4.47 (0.66)	4.61(0.51)
Part-time special education	3.23 (0.83)	3.09 (0.83)	3.46 (1.05)
Homework support	2.92 (1.04)	3.32 (1.04)	3.69 (0.85)
Complementary oral examinations	3.85 (0.80)	3.79 (0.94)	4.15 (0.99)
Co-teaching	2.23 (1.01)	3.21 (1.04)	3.39 (1.33)
Remedial education	3.85 (0.69)	3.71 (0.76)	3.77 (0.83)

Table 2. Items (translated from Swedish), correlation, mean and standard deviation for the teacher efficacy scale.

	1	2	3	4	5	6	7	8
1. I have enough knowledge about difficulties in mathematics and know what to do.	1.000							
2. I have a feeling of hopelessness.	**-.466**	1.000						
3. I often ask for advice.	-.247	.151	1.000					
4. It is challenging, but I manage well.	-.002	.023	.031	1.000				
5. I seldom teach low-achieving students myself; the mathematics/special education teacher takes care of them.	**-.404**	**.448**	.141	.0138	1.000			
6. I feel doubtful, but with help I manage.	**-.601**	**.488**	.197	-.125	**.571**	1.000		
7. I get too little help from colleagues (extra resources).	-.113	.080	.084	.299	.008	.135	1.000	
8. I need more knowledge about difficulties in mathematics.	**-.520**	**.458**	.144	.064	**.324**	**.582**	.206	1.000
M	3.07	1.82	2.60	2.81	1.97	1.95	2.40	2.31
SD	0.58	0.67	0.63	0.58	0.88	0.80	0.79	0.82

The total scores from items 1,2,5,6 and 8 were included in the teacher efficacy beliefs variable.