

JYX



This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Kiviniemi, Tiina; Nuora, Piia

Title: Evaluation of a Chemistry Concept Inventory for general chemistry students at Finnish university

Year: 2020

Version: Published version

Copyright: © Authors, 2020

Rights: CC BY 3.0

Rights url: <https://creativecommons.org/licenses/by/3.0/>

Please cite the original version:

Kiviniemi, T., & Nuora, P. (2020). Evaluation of a Chemistry Concept Inventory for general chemistry students at Finnish university. *NorDiNa*, 16(2), 138-148.

<https://doi.org/10.5617/nordina.7014>

Tiina Kiviniemi is an experienced University Teacher at the Department of Chemistry, University of Jyväskylä, with a background of pedagogical studies and a PhD in Physical Chemistry. She teaches and gives study counselling to chemistry students at bachelor level. Her research is focused mainly on understanding and developing teaching and learning processes in bachelor level University Chemistry Education.

Piia Nuora is a science teacher at Kannonkoski Comprehensive School. She received MA (Educ.) in the year 2014 in Faculty of Education and her PhD in the year 2016 in Chemistry Education, University of Jyväskylä, Finland. Her research is mainly focused on Chemistry Education, out-of-school learning in STEM subjects and drawing research.

TIINA KIVINIEMI

Department of Chemistry, University of Jyväskylä, Finland
tiina.t.kiviniemi@jyu.fi

PIIA NUORA

Kannonkoski Comprehensive School, Finland
piia.nuora@outlook.com

Evaluation of a Chemistry Concept Inventory for general chemistry students at Finnish university

Abstract

A chemistry concept inventory (Chemical Concept Inventory 3.0/CCI 3.0), previously developed for use in Norwegian universities, was tested and evaluated for use in a Finnish university setting. The test, designed to evaluate student knowledge and learning of chemistry concepts, was administered as both pre- and posttest in first year general chemistry courses at the University of Jyväskylä. The results were evaluated using different statistical tests, focusing both on individual item analysis and the entire test. Some individual questions were found to be not discriminating or reliable enough or too difficult, yet the results, as a whole, indicate that the concept inventory is a reliable and discriminating tool that can be used in the Finnish university context.

INTRODUCTION

In recent decades, new theories of learning and associated methods of teaching have appeared in the fields of science, technology, engineering, and mathematics (STEM). New assessment tools are being created to measure students' conceptual knowledge to compare the differences in learning outcomes between new and old methods, and also to understand what background limitations and preconceptions students have when entering a class (Krause, Birk, Bauer, Jenkins & Pavelich, 2004). According

to Nyachwaya et al. (2011), researchers have developed concept inventories of ideas that are central to the discipline. These inventories have been used to assess students' conceptual understanding. For example, Mulford and Robinson (2002) have developed the Chemistry Concepts Inventory (CCI), a 22-item multiple-choice inventory. CCI was meant to measure students' conceptual understanding of common topics taken in general chemistry class. It was also meant to measure alternate conceptions about these topics (Barbera, 2013; Fajardo & Bacarrisas, 2017). Furthermore, Mulford and Robinson (2002) have also examined how these alternate conceptions changed after one semester in such a course.

Conceptual understanding of science is a complicated phenomenon. It comprises declarative knowledge, procedural knowledge and conditional knowledge. Procedural knowledge includes models, rules, algorithms, while conditional knowledge includes the understanding of when to employ procedural knowledge. It also includes an understanding of why it is important to do so (Niewswandt, 2007). As an assessment tool, CCI is seen as a useful tool for comparing different methods and to measure students' conceptual understanding (Kiviniemi, Eggen, Persson, Hafskjold & Jacobsen, 2017).

PURPOSE OF THE STUDY

The aim of this research project is to find out whether the CCI 3.0 test, developed at the Norwegian University of Science and Technology (NTNU), can be used to evaluate student knowledge and learning of chemistry concepts in a Finnish university setting (see also Kiviniemi et al., 2017; Eggen, Jacobsen, Hafskjold, & Persson, 2017). Using the test for analysis purposes requires it to discriminate between different students and to be statistically reliable. After this is confirmed, the test results can be used to analyze students' knowledge of chemical concepts, their possible misconceptions, and, when using the test both pre- and post-instruction, changes in students' learning outcomes.

The study described in this paper aims to answer the following questions: (1) Is the CCI 3.0 test statistically discriminating and reliable enough to be used for Finnish university students? (2) Are there any statistically problematic test questions and is it possible to suggest reasons for their failure for use in developing the test further?

RESEARCH METHODS

Chemistry Concept Inventory (CCI 3.0)

The chemistry concept inventory used in this study has been developed at NTNU for use in general chemistry courses in Norwegian universities and upper secondary schools in Norway (Eggen, Jacobsen, Hafskjold, & Persson, 2017). The purpose of the inventory is to map students' understanding of concepts in general chemistry, and to use the results from the test as a tool for evaluating learning and teaching in general chemistry courses. The CCI test was developed gradually at NTNU based on existing chemistry concept inventories and by adding questions and concepts based on the personal experience of the developers as well as on the literature references (Krause et al., 2004; Mulford & Robinson, 2002). Because the test is intended for use in evaluating chemistry concepts held by first year chemistry students, it was mainly developed to cover the main topics introduced in undergraduate chemistry courses rather than the majority of all chemical concepts.

The version used in this study, CCI 3.0, is a multiple-choice test with 40 questions, covering the main topics of the first-year university general chemistry courses, including chemical bonding, atomic and molecular structure, intermolecular forces, different chemical equilibria, electrochemistry and thermodynamics. A more detailed listing of the topics can be found in Eggen et al. (2017). The test has an estimated completion time of 45 minutes and was previously tested at NTNU as a posttest in a general chemistry course during the 2015 spring semester (Eggen et al., 2017).

For use in this study, the CCI 3.0 was translated into Finnish and modified into an electronic form on the Moodle learning platform (Moodle.org). The topics of the questions in CCI 3.0 (Eggen et al., 2017) were found to match the contents of the first two autumn semester general chemistry courses at the University of Jyväskylä (JYU), so there was no need to modify the contents of the test (University of Jyväskylä, 2016). The order of the multiple-choice options and the pictures included were equivalent to the test at NTNU. The test was administered at JYU in two successive, first-year general chemistry courses in the autumn semesters of 2016 and 2017.

Data collection

The pretest was part of the Basics of Chemistry 1 course exercises and was compulsory if a student wanted to include exercise points in the course grading. The test answers did not affect exercise points. The only thing controlled in grading was if the student had answered or not. Students filled in the questionnaire during the first week of the course. The questionnaire was available in the course learning environment (Moodle), which was open only for students registered in the course. Before answering, how the data would be used was communicated to the students, and they could choose if they allowed their course information and answers to be combined and used as data for the study. Students were not supervised while taking the test, but the instructions emphasize, that using external materials was not needed, and that it was important to answer according to their current knowledge because the purpose of the test was to find out what they really know. According to test developers' wishes, the students were not given the answers for the questions nor their personal results after the test, and the course teachers did not go through the answers during the course. This was done to protect the questions from being distributed among other students and to ensure that students answering the same questions on the posttest had not memorized the correct answers but rather answered according to their current knowledge.

The data collection parameters and student backgrounds are summarized in Table 1. Not all students who were registered for the course completed the test, and some of the answer sets were removed from the data according to the following criteria: (1) the student did not give permission to use their answers in the study (denied permission or left the question unanswered), (2) the student answered less than half (20) of the CCI questions, or (3) the student selected the same letter/number option for more than half of the CCI questions. The third criterion was used to remove answer sets where a student had probably not read the questions but only clicked the same answer throughout the test.

The posttest was administered along the same guidelines as the pretest, but during the last week of the second part of the Basics of Chemistry (Basics of Chemistry 2 course) in the first autumn of chemistry studies. The link for the posttest was also sent to students who were not attending the second course but had attended the first one. Answers usable for analysis were selected with the same criteria as for the pretest.

The test groups in the data used for analysis consisted mainly of students from the Faculty of Mathematics and Science, majoring in biological and environmental sciences, chemistry, and physics. The rest of the students were from varying disciplines such as mathematics, information technology, and a variety of non-science subjects. The groups consisted mainly of first-year students, but there were also some students with chemistry as a minor who had already studied for two or more years at the university. (See Table 1)

Table 1. Summary of the data collection and student background information for the test groups.

		Pretest		Posttest	
		2016	2017	2016	2017
N	Students registered	187	178	161	156
	Answered the test	172	174	121	94
	Usable in analysis*	141	147	79	76
Major subject (%)	Bio- and env. sciences	49	53	46	29
	Chemistry	23	29	33	43
	Physics	16	4	14	12
	Other**	12	14	7	16
Study year (%)	First	71	75	72	62
	Second	16	10	15	20
	Older	13	15	13	18

*See text (Data collection) for the criteria used

**Mathematics, Information technology, and a variety of non-science subjects

RESULTS

The CCI test results were analyzed using different statistical tests, focusing on individual items and on the test as a whole. The analysis was done separately for both pre- and posttest to see possible differences when using the CCI before or after instruction, and whether the test can be used for both purposes. To obtain comparable results, the analysis methods described below were the same that were used for the CCI 3.0 at NTNU (Eggen et al., 2017).

Concept inventory as a whole

The reliability and discriminatory power of the test were evaluated using the Kuder-Richardson test reliability index (r_{test}) and Ferguson's delta (δ) (Kline, 1986; Kuder & Richardson, 1937). The Kuder-Richardson reliability index evaluates the reliability of the test by comparing individual test items, considering each item as a single parallel test. r_{test} can be calculated for multiple choice tests as (Kuder & Richardson, 1937):

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

where M is the number of items in the test, σ_x is the standard deviation of the total test score, and P is the difficulty index for each item, calculated as the ratio of the number of correct answers N_1 to the total number of answers N :

$$P = \frac{N_1}{N} \quad (2)$$

The test can be considered reliable for assessing individuals if the value of the Kuder-Richardson reliability index is greater than 0.8 (Doran, 1980).

The other test statistic for the test, Ferguson's delta, measures the discriminatory power of the test based on the students' score distribution (Kline, 1986):

$$\delta = \frac{N^2 - \sum f_i^2}{N^2 - (N^2 / (M + 1))}, \quad (3)$$

where f_i is the frequency of cases with the same score, and N is the number of students taking the test. The test can be considered to have good discrimination if the value of Ferguson's delta is greater than 0.9 (Kline, 1986). The results for these two statistical tests for the pre- and posttest in JYU are shown in Table 2 together with previous results from NTNU (Eggen et al., 2017).

Table 2. Comparison of statistical test results for CCI 3.0. N = number of students, r_{test} = Kuder-Richardson test reliability, and δ = Ferguson's delta.

	JYU Pre 2016	JYU Post 2016	JYU Pre 2017	JYU Post 2017	NTNU Post 2015 ^a	Recommended value ^b
N	141	79	147	76	60	
r_{test}	0.85	0.85	0.84	0.86	0.88	> 0.8
δ	0.98	0.98	0.98	0.97	0.98	> 0.9

a) From Eggen et al. (2017)

b) Doran (1980), Kline (1986)

As can be seen in Table 2, both pre- and posttest in JYU pass the statistical tests, and the values are comparable to the results from NTNU. Thus, the CCI 3.0 can be considered a reliable and discriminating tool that can be used to map students' knowledge of chemical concepts in general chemistry courses at the University of Jyväskylä. Considering the similar contents of general chemistry courses and the similarity of student backgrounds in Finnish universities, it appears possible that the test could also be used in different Finnish universities with similar results.

Individual questions

The individual test questions were evaluated using three different statistical tests: item difficulty index P , item discrimination index D , and item point biserial coefficient r_{pbc} . Item difficulty index (Equation (2)) is simply the ratio of the number of correct answers to the total number of answers for a certain item. Thus, the greater the value of P , the easier the question is for the students. A question or a test that is too easy or difficult will not properly discriminate between students. A reasonable value for the difficulty index should be between 0.3 to 0.9 for individual questions, where 0.5 would be the optimum. The average value of the item difficulty index for the whole test should be around 0.5. (Doran, 1980)

The item discrimination index D measures how well an individual test item discriminates between high- and low scoring students. This is based on dividing the students into high- and low-scoring groups and calculating the number of correct answers in each group. The division can be made in different ways, but here we used a 25 % / 25 % division, meaning the high-scoring group consisted of the best 25 % of students and the low scoring group of the 25 % that had the lowest scores. Using this division, the item discrimination index can be calculated as follows (Doran, 1980):

$$D = \frac{N_H - N_L}{N/4}, \quad (4)$$

Evaluation of a Chemistry Concept Inventory for general chemistry students

where N_H and N_L are the numbers of correct answers in the high-scoring and low-scoring groups, respectively, and N is the total number of students who answered the question. A high value of D (close to +1) suggests that most of the high scoring students, and almost none of the low scoring group, answered correctly. A negative value suggests that low-scoring students answered the question correctly more frequently than did the high scoring students, indicating poor discrimination. A question can be considered to have good discrimination if $D \geq 0.3$, and questions with a negative value should be modified or eliminated from the test. (Doran, 1980)

The third statistical test used for the individual items is the item point biserial coefficient r_{pbc} . It measures the reliability of an individual item by correlating the total score and the score on an individual item. If r_{pbc} has a positive value, a student with a high total score is more likely to answer the item correctly than is a student with a low score. The point biserial coefficient for an item can be calculated using the item difficulty index P (Eq. (2)) and the standard deviation of the total test score σ_x (Kline, 1986):

$$r_{pbc} = \frac{\bar{x}_1 - \bar{x}_0}{\sigma_x} \sqrt{\frac{P}{1-P}}, \quad (5)$$

where \bar{x}_1 is the average total score for those who answered an item correctly, and \bar{x}_0 is the average total score for all. The test as a whole should have an average item point biserial coefficient value larger than 0.2, but a few individual questions with a value lower than that can be accepted in the test (Kline, 1986).

The average results with minimum and maximum values for the individual item statistics are presented in Table 3. The average values for pre- and posttests in JYU were within recommended limits and thus acceptable. The average item discrimination index and point biserial coefficient were slightly lower in JYU than they were in NTNU, but the averages were still well above the limits. The average value of the item difficulty index was around 0.5 for both pre- and posttests, which can be considered a proper value.

Table 3. Comparison of statistical test results of individual questions (mean [min, max]) for CCI 3.0. N = number of students, P = item difficulty index, D = item discrimination index, r_{pbc} = item point biserial coefficient.

	JYU Pre 2016	JYU Post 2016	JYU Pre 2017	JYU Post 2017	NTNU Post 2015 ^a	Recommended value ^b
N	141	79	147	76	60	
P	0.46 [0.19, 0.79]	0.57 [0.14, 0.90]	0.46 [0.22, 0.74]	0.62 [0.25, 0.91]	0.66 [0.33, 0.92]	0.5 [0.3, 0.9]
D	0.44 [-0.03, 0.72]	0.42 [0.05, 0.73]	0.42 [-0.15, 0.72]	0.43 [-0.05, 0.90]	0.45 [0.13, 0.73]	≥ 0.3
r_{pbc}	0.38 [0.06, 0.56]	0.37 [0.10, 0.63]	0.36 [-0.13, 0.60]	0.39 [0.08, 0.68]	0.41 [0.15, 0.56]	> 0.2

a) From Eggen & al. (2017)

b) Doran (1980), Kline (1986)

When comparing the results from pre- and posttests in JYU, it is apparent that the test was easier as a posttest than it was as a pretest, which is expected in a case where the posttest is preceded by instruction on the topics of the test. Even though the pretest appears to be somewhat difficult, the average difficulty indexes of the posttests at both JYU and NTNU indicate that the test cannot be made much easier. An easier test might lead to ceiling effects when using the same test as a posttest (Persson, 2015). Thus, a pretest that is slightly “too difficult” may be more optimal, if the test is still discriminating and reliable enough. The NTNU 2015 student group was assumed to be high achieving (Eggen et al., 2017), which may explain the notably higher value of the item difficulty index compared to JYU students.

However, when looking more closely at the statistical test values for individual test items, there are some questions that appear to be problematic. In the JYU 2016 pretest, the item difficulty index was smaller than 0.3 for nine questions, the item discrimination index was below 0.3 for seven questions and negative for one of them, and the point biserial coefficient was less than 0.2 for three questions. In the 2017 pretest, the item difficulty index was smaller than 0.3 for eight questions, item discrimination index was below 0.3 for six questions and negative for one of them, and point biserial coefficient was less than 0.2 for four questions. In both years’ groups, most of these values differed from the recommended limits only slightly, but there were two questions that failed all three statistical tests in both years, and one that failed both D and r_{pb} tests, that is, they did not discriminate well between students. The questions failing statistical tests were the same at JYU in 2016 as well as in 2017.

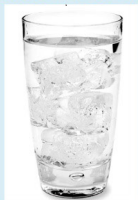
In the JYU posttest 2016, only two questions had difficulty indexes significantly below 0.3, and in 2017 only one did, indicating that the questions were easier for the students after instruction in general chemistry courses. The item discrimination index in the JYU 2016 posttest was less than recommended for nine questions, and significantly below the limit for seven questions, but had no negative values. In 2017, the item discrimination index was below recommendation only for six questions, of which one had a negative value. In both years, the point biserial coefficient for posttest was below the recommendation for three questions, but significantly only for one of them. In both JYU posttests, only one of the questions failed all three statistical tests, with similar values as on the pretest.

It is apparent that a question failing all three statistical tests should either be removed from the test or modified so that it can discriminate between students. However, simply removing the question may not be the optimal choice because it would also remove a chemical concept from the list of topics to be tested. Further, a more detailed analysis of the question can also help in developing teaching on the general chemistry courses. A careful analysis of the question, considering possible misconceptions, can offer educators valuable insights on topics that are difficult for students. Here, we consider the two questions that failed all three statistical tests, one in pretest and the other in both pre- and posttests.

Analysis of the statistically most problematic questions

The question that failed all three statistical tests only in the pretest (in 2016 and 2017) is presented in Figure 1. In the pretest, only 74/89 % (2016/2017) of students answered this question when the average answer percentage was 95 %, which indicates that the question was found to be difficult or not understood by the students. The difficulty index for this question in the pretest was 0.22/0.24, which is close to the value that would result from pure guessing (0.2, with 5 multiple choice options, of which one is correct). In the pretest, this question also had the smallest item discrimination index (-0.03/-0.15) and point biserial coefficient (0.06/-0.13) of all the questions, showing that the question did not discriminate between high-scoring and low-scoring students. This leads to the conclusion that most of the students who answered the question were probably guessing.

Melting of ice can be described with chemical equation: $\text{H}_2\text{O}(\text{s}) \rightarrow \text{H}_2\text{O}(\text{l})$. Which of the following is correct, when this reaction occurs at room temperature in a glass filled with ice water?



The glass in the picture contains ice cubes and water. The system is not in equilibrium, and the ice is melting.

Select one:

- A. ΔH and ΔS have opposite signs
- B. $\Delta G = 0$
- C. Both ΔH and ΔS are greater than zero
- D. $\Delta G > 0$
- E. If the temperature decreases, both ΔH and ΔS increase

Figure 1. Question 34 from CCI 3.0, translated into English by the authors. This question failed all three statistical tests in the JYU pretest but was statistically acceptable for the posttest.

The explanation for the difficulty of this question lies most probably in the chemistry education background of the students. To be able to answer this question, a student must be familiar with the concepts of enthalpy, entropy and Gibbs free energy, know the connection between them and the spontaneity of a reaction, and recognize the symbols used. However, in Finnish national core curriculum for general upper secondary schools (Finnish National Agency for Education, 2004; Finnish National Agency for Education, 2016), these topics are not required to be covered except for enthalpy changes in chemical reactions. The students, whose background in chemistry comes mainly from upper secondary education, most probably did not know these concepts and symbols, and were thus not able to answer this question. However, after one semester of instruction in general chemistry, the situation has changed, as indicated by the posttest. The question in the JYU posttest was still difficult ($P = 0.23/0.31$), but most of the students (94/99 %) answered it, and the question discriminated properly between high- and low-scoring students ($D = 0.41/0.38$ and $r_{pbc} = 0.33/0.36$). The analysis of this question clearly shows the effect of instruction, so it can be used in pre- and posttest setting. But if the pretest results are to be used without connection to the posttest, discarding the answers to this question from the analysis of the test would increase the reliability and discrimination of the CCI 3.0 pretest results.

One of the CCI 3.0 questions failed all three statistical tests in both the pre- and posttest at JYU. The question, featuring electrochemistry, is presented in Figure 2. To answer this question correctly, a student had to recognize the terminology of an electrochemical cell, and then understand the process (reduction) that is taking place. The discriminatory power of this question was poor in both pretests ($D = 0.17/0.03$ and $r_{pbc} = 0.19/0.04$ in 2016/2017) and posttests ($D = 0.14/-0.05$ and $r_{pbc} = 0.19/0.08$), and surprisingly, the item difficulty index was smaller in the posttest ($P = 0.14/0.25$) than it was in the pretest ($P = 0.22/0.29$). This topic is included in Finnish national core curriculum for upper secondary schools, so the students should be able to answer this question if they have studied elective chemistry courses at the upper secondary level (Finnish National Agency for Education, 2004; Finnish National Agency for Education, 2016). Obviously, this is not the case among JYU students, and instruction on this topic at the university seems to have no or even negative effect on the students' understanding of the concept. The statistical test results for this question suggest that as such, when

considering discrimination and reliability, the question is not suitable for CCI 3.0, and should either be discarded from the test analysis or modified before using it. To understand the problems with this question more thoroughly, additional investigation, more detailed analysis of the answer profiles, and/or testing different types of questions on the same topic is currently under consideration.

Consider a copper cathode with half-reaction: $\text{Cu}^{2+}(\text{aq}) + 2 \text{e}^{-} \rightarrow \text{Cu}(\text{s})$. The copper ions in solution gain electrons from _____, forming metallic copper.

Select one:

- A. solution
- B. cathode
- C. anode
- D. salt bridge
- E. external circuit

Figure 2. Question 19 from CCI 3.0, translated into English by the authors. This question failed all three statistical tests in the JYU pretests and posttests.

CONCLUSIONS

The main purpose of this study was to find out whether the CCI 3.0 test is statistically discriminating and reliable enough to be used in a Finnish university setting. Statistical tests were used to evaluate the difficulty, discriminatory power, and reliability of the individual questions, and of the CCI 3.0 as a whole. The test was found to be discriminating and reliable enough to be used for assessing conceptual changes at the general chemistry level in a Finnish university setting (e.g. Eggen et al., 2017; Kiviniemi et al., 2017).

According to the statistical tests of the individual questions, most of the questions are within acceptable limits for proper difficulty and discrimination, so they can be used for testing the chemistry concepts of general chemistry students. Some of the individual questions were found to be rather difficult when given as a pretest, but in the posttest these questions became easier, as expected when instruction is given on the test topics between the pre- and posttests. Because the difficulty of the test is always dependent on the group taking the test, the difficulty index is bound to vary from test to test. The results for the difficulty index in different groups in this study are all within a reasonable range, so there is no need to adjust the test difficulty.

A couple of individual questions in CCI 3.0 were found to not properly discriminate between high- and low-scoring students at JYU. These questions were the same both years, indicating a need to analyze them further. The question discussing thermodynamics and the spontaneity of a reaction failed to discriminate between students only in the pretest, which is probably due to the fact that the topic of the question is not fully covered in chemistry studies preceding university. Thus, when analyzing only the pretest results, this question should probably be left out of the analysis, at least at Finnish universities, to improve discrimination. However, in the posttest, or when analyzing the changes in students' knowledge of chemistry concepts, the question can be used to find out learning gains in the general chemistry courses. For the electrochemistry question that failed to discriminate between students in both pre- and posttest, there was no clear explanation found. As such, this individual question should be left out when analyzing the pre- and posttests, or totally discarded from the CCI 3.0. However, a better approach would be to develop the question further so that the weight of electrochemistry as a topic would not be changed in the test. Further research on this question is under consideration so that the question can be developed to be suitable for testing electrochemistry concepts among students.

Overall, students need help in developing their conceptual understanding of the specific basic concepts and principles of general chemistry (Fajardo & Bacarrisas, 2017). Using a tool such as a CCI to test chemistry concepts before and after instruction, it is possible for educators to evaluate whether the support and teaching that the students are receiving has the desired effect.

In our study, the CCI 3.0 test was found to be statistically suitable for use among general chemistry students in a Finnish university setting. Additional research on the test is however still needed to resolve the issues with a few problematic questions. The use of CCI 3.0 has been continued at JYU in the autumn semesters of 2018 and 2019 to gather more data and to follow the effects of some pedagogical changes made in the general chemistry courses. In the future, this research will be used to develop a new, even more suitable version of the CCI test, the results of which can be used to gain insight into the conceptual knowledge of chemistry and its changes among general chemistry students in Finland.

ACKNOWLEDGEMENTS

The authors wish to thank the CCI development team at NTNU for proposing collaboration and providing the CCI 3.0 test for our use in this research. Dr. Jonas Persson is also warmly thanked for his help in the statistical analysis of the data.

REFERENCES

- Barbera, J. (2013). A psychometric analysis of the chemical concepts inventory. *Journal of Chemical Education*, 90(5), 546–553. doi:10.1021/ed3004353
- Doran, R. L. (1980). Basic measurement and evaluation of science instruction. National Science Teachers Association, 1742 Connecticut Ave., N.W., Washington, DC 20009.
- Eggen, P., Jacobsen, E. E., Hafskjold, B., & Persson, J. (2017). Development of an inventory for alternative conceptions among students in chemistry. *LUMAT: International Journal on Math, Science and Technology Education*, 5(1), 1-11. doi:10.31129/LUMAT.5.1.115
- Fajardo, M. T. M., & Bacarrisas, P. G. (2017). First-year college students' knowledge in chemistry: Is it adequate? *American Journal of Educational Research*, 5(10), 1039–1043. doi:10.12691/education-5-10-5
- Finnish National Agency for Education. (2004). National core curriculum for general upper secondary education 2003 (1st ed.) Finnish National Agency for Education.
- Finnish National Agency for Education. (2016). National core curriculum for general upper secondary schools 2015 (1st ed.) Finnish National Agency for Education.
- Kiviniemi, T., Eggen, P., Persson, J., Hafskjold, B., & Jacobsen, E. (2017). Development of a chemistry concept inventory for general chemistry students at Norwegian and Finnish Universities. In Auður Pálsdóttir (Ed.), *Proceedings of the 12th Nordic Research Symposium on Science Education 2017* (pp. 73-79).
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. London: Methuen.
- Krause, S., Birk, J., Bauer, R., Jenkins, B., & Pavelich, M. J. (2004) Development, testing, and application of chemistry concept inventory. ASEE/IEEE 34th Annual Frontiers in Education Conference. Savannah. doi:10.1109/FIE.2004.1408473
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160. doi:10.1007/BF02288391
- Moodle.org. (n.d.). Open-source learning platform. <https://moodle.org/>
- Mulford, D. R., & Robinson, W. R. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of Chemical Education*, 79(6), 739–744. doi:10.1021/ed079p739
- Nieswandt, M. (2007). Student affect and conceptual understanding in learning chemistry. *Journal of Research in Science Teaching*, 44(7), 908–937. doi:10.1002/tea.20169

- Nyachwaya, J. M., Mohamed, A.-R., Roehrig, G. H., Wood, N. B., Kern, A. L., & Schneider, J. L. 2011. The development of an open-ended drawing tool: An alternative diagnostic tool for assessing students' understanding of the particulate nature of matter. *Chemistry Education Research and Practice*, 12, 121-132. doi:10.1039/C1RP90017J
- Persson, J. R. (2015). Evaluating the force concept inventory for different student groups at the Norwegian University of Science and technology. Retrieved May 13, 2017, from <https://arxiv.org/abs/1504.06099>
- University of Jyväskylä, Faculty of Mathematics and Science. (2016). Opinto-opas 2016-2017. (In English: Study Guide 2016-2017) (pp. 266-267). Retrieved February 28, 2020, from University of Jyväskylä, Faculty of Mathematics and Science web site: https://www.jyu.fi/science/fi/ohjeita-opiskelijalle/tutkinto-ohjelmat/vanhat-opetussuunnitelmat/mltk_opinto-opas_lv2016-2017.pdf