Development of a Chemistry Concept Inventory for General Chemistry Students at Norwegian and Finnish Universities

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

Per-Odd Eggen and Jonas Persson
Department of Teacher education, Norwegian University of Science and Technology, Trondheim, Norway

Bjørn Hafskjold and Elisabeth Jacobsen
Department of Chemistry, Norwegian University of Science and Technology, Trondheim, Norway

Abstract A Chemistry concept inventory has been developed for assessing students' learning and identifying alternative conceptions that students may have in general chemistry. The inventory aims at functioning as a tool for adjusting teaching practices in chemistry and is mainly aimed at assessing students' learning during general chemistry courses. The inventory was administered as a post-test in a general chemistry course at the Norwegian University of Science and Technology (NTNU) in spring 2015, and evaluated using different statistical tests, focusing both on item analysis and the test as a whole. The results from this analysis indicated that the concept inventory is a reliable and discriminating tool as a post-test in the Norwegian context. Here, we present results from the statistical analysis of the test, when administered as both pre- and post-test in general chemistry courses at both University of Jyväskylä, Finland (JYU) and at NTNU in autumn 2016. The comparison of the results of the statistical analysis from these tests show that the concept inventory developed in NTNU could be used as a tool for investigating students' learning in both Norwegian and Finnish university contexts.

Keywords General chemistry, Higher education, Concept inventory, Misconception, Test development

1. Introduction

1.1 Background

Learning natural sciences can be described as a process where the learner develops and revises concepts of how the nature works based on intuitive ideas, observations and theoretical models. The continuous process of developing and revising these concepts may be induced by formal education or informal, individual observations. In both cases, the concepts already held will be the starting point (Smith et al. 1994).

A Chemistry concept inventory (CCI) was developed for use in general chemistry courses at Nordic universities as well as in upper secondary schools (Eggen et al. 2017). The inventory is aimed to serve two purposes: i) to map students' understanding of concepts and phenomena in chemistry and ii) as an independent tool for evaluation of learning activities. 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.

CCI is used as a pre-test in the beginning of first university chemistry course to gain information on students' chemistry conceptions at the beginning of their university study. These conceptions are assumed to be either acquired in the school subject (such as the mole concept) or in daily life (such as the belief that matter disappears when something burns). One purpose for using the CCI as a pre-test may be to obtain information about which concepts students hold after completing the secondary school chemistry education. Another purpose with the pre-test may be to map students' starting point before the university courses. Knowledge about student's understanding of concepts may be used for individual guidance, for designing a pre-course or for adjusting the first-year chemistry education to specific student groups. If the CCI results indicate unfortunate conceptions, it may be possible to address specific parts of the course content that need to be improved. However, the observed problems may also stem from other related issues such as focus on exam rather than understanding the course content, or a suboptimal study strategy. These issues are not directly assessed in the CCI but their effect can be evaluated combining the data with for example learning attitudes test such as CLASS-Chemistry (Adams et al. 2008).

On the other hand, CCI as a post-test shows students' conceptions after the university course. The results from the post-test can be used as an indication of specific areas where students have problems with their conceptual understanding, which will help in reforming courses to address these problems. The results from different universities provide a possible tool for comparing study programs and identify strengths and weaknesses of the courses given. Since the knowledge students possess must be assumed to be a result of both primary, secondary and university education, one could aim at addressing all these levels to improve the overall outcome of the chemistry study.

In the Nordic countries, there has not been any study published regarding student's conceptual knowledge in chemistry. There are some studies published from the USA. Most of them use the concept inventories which we used as templates for the initial bank of questions (Krause et al. 2004, Mulford & Robinson 2002).

1.2 Theoretical framework

Concept inventories are aimed at describing the concepts held by students, and are known to be used in the fields of Science, Technology, Engineering and Mathematics (STEM). Concepts are developed from early age, and children form intuitive ideas of natural phenomena. Concepts not consistent with the established consensus are sometimes called misconceptions (Smith et al. 1994) or alternative conceptions. Several studies have mapped chemistry students' concepts based on beliefs and described alternative concepts concerning chemical phenomena. (Bowen & Bunce 1997, Gabel & Bunce 1994, Krajcik 1991, Nakhleh 1992, Stavy 1991, Stavy 1995, Wandersee et al. 1994).

A concept inventory consists normally of a series of multiple-choice questions, based on qualitative problems. It aims to measure deep understanding and conceptual knowledge rather than a student's ability to solve problems. For example, the Force Concept Inventory (FCI) for use in physics education was created by Hestenes et al. (1992) and is quite commonly applied in teaching and research. Chemical concept inventories are also developed (Krause et al. 2004, Mulford & Robinson 2002), although they do not seem to be commonly used in studies in the Nordic countries.

Tools for mapping and analyzing students' conceptions may help the teacher to adjust their practice to facilitate deeper understanding. The students are assumed to build conceptual understanding from their present conceptions, and they might find it difficult to accept new information that does not fit into their existing beliefs. This conflict between student's established views and the new information can be disbelieved and rejected, or accepted with minor or more extensive changes in the student's conception.

Compared to teacher-centered instruction, a more student-centered model of education using more hands-on and inquiry-based approach, is believed to increase the student's knowledge and conceptual understanding of a subject (Taber, 2009). Assessment tools, such as the CCI, can be useful for comparing different methods and to measure the students' conceptual understanding as well as to understand what conceptions and background limitations the students have when entering a class.

When developing assessment tools, such as Chemistry concept inventory, the tool needs to be evaluated considering the validity and reliability of the test, before drawing further conclusions from the test results. The validity, i.e. whether the test covers a proper range of topics related to the subject, is usually assessed by expert consensus. The reliability of the test includes consistency and discriminatory power. If the test is consistent, a student would get the same score when taking the test again, assuming no change in the student's thinking. The discriminatory power of the test indicates if the test can separate between students with different levels of understanding on the subject. These aspects of reliability can both be evaluated using statistical analysis of the test results. (Arjoon et al. 2013, Kline 1986).

2. Method

The CCI test was developed at NTNU based on existing chemistry concept inventories and by adding questions on concepts based on personal experience of the developers and on literature references (Krause et al. 2004, Mulford & Robinson 2002). The concept inventory CCI 3.0 analyzed here, consisting of 40 questions, was developed gradually from a larger set of questions by Eggen et al. (2017), evaluating the discriminatory power and difficulty of the questions and taking into account the time needed for completing the test. For the current CCI 3.0 test, the recommended answer time is 40-60 minutes. As the aim was to use the test to evaluate chemistry concepts held by first year chemistry students, the test was mainly developed to cover main topics introduced in undergraduate chemistry courses rather than the majority of all chemical concepts. The content validity of the test was evaluated by expert assessment in NTNU. CCI 3.0 was administered at NTNU as a post-test in general chemistry course during the 2015 spring semester and analysed using statistical tests for reliability and discrimination power (Eggen et al. 2017).

In a following study, the applicability of CCI 3.0, both as a pre- and post-test was evaluated in different university settings. The test was administered as both preand post-test in general chemistry courses in JYU and NTNU during the 2016 autumn semester. In JYU, the pre-test was administered in the first week of the first university general chemistry course in September 2016, and the post-test during the last week of the second general chemistry course in December 2016. The test was administered electronically in the course's Moodle learning platform. At NTNU the tests were administered on paper; the pre-test was given the first week and the post-test was given the last week of the semester. The students at NTNU had only about 30 minutes in the end of a lecture to answer the CCI, which turned out to be a too short time. This can be seen in the number of unfinished tests – only few students in this group answered all the questions. Both in JYU and NTNU the student groups were mainly first year students, including both students with chemistry as a major subject and students majoring in different fields of science (mostly physics and biosciences).

3. Results and discussion

The aim of this study was to statistically evaluate the consistency and discriminatory power of CCI 3.0. The test results were analysed using five statistical tests: three focusing on individual items (item difficulty index P, item discrimination index D, item point biserial coefficient r_{pbc}) and two on the test as a whole (Kuder-Richardson test reliability r_{test} and Ferguson's delta δ). (Kline 1986, Kuder & Richardson 1937). An analysis of individual questions considering the answer distributions and misconceptions is outside the scope of this paper and will be presented in an extensive analysis in later publications.

The results from statistical tests are presented in Table 1. They were found to be similar in both universities, with some slight differences as expected when testing different student groups. For the NTNU 2016 pre-test, in which the answer time was too limited and a large number of students couldn't finish the test, the mean values for two out of three statistical tests for individual questions fail to fit in the recommended limits (Doran 1980, Kline 1986). However, in other samples, mean values of the statistical tests for individual questions are within the recommendations. The tests for CCI 3.0 as a whole, Kuder-Richardson reliability index r_{test}, and Ferguson's δ, which measures discriminatory power of the whole test, are well within recommendations for all tests. Thus, in an overall picture the CCI 3.0 appears to be a reliable and discriminating tool within the contexts it has been used.

Table 1. Comparison of statistical test results (*mean* [*min, max*]) for CCI 3.0 in different groups. N = number of students, P = item difficulty index, D = item discrimination index, r_{pbc} = item point biserial coefficient, r_{test} = Kuder-Richardson test reliability, and δ = Ferguson's delta.

	JYU Pretest 2016	JYU Posttest 2016	NTNU Pretest 2016[a]	NTNU Posttest 2016[a]	NTNU Posttest 2015[b]	Recommended value[c]
N	141	79	42	71	60	
P	0.46 [0.19, 0.79]	0.57 [0.14, 0.90]	0.37 [0.07, 0.78]	0.55 [0.22, 0.84]	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]	0.23 [-0.23, 0.77]	0.47 [-0.11, 0.84]	0.45 [0.13, 0.73]	≥ 0.3
r_{pbc}	0.38 [0.06, 0.56]	0.37 [0.10, 0.63]	0.34 [-0.21, 0.68]	0.42 [-0.03, 0.65]	0.41 [0.15, 0.56]	> 0.2
r_{test}	0.85	0.85	0.80	0.88	0.88	> 0.8
δ	0.98	0.98	0.95	0.98	0.98	> 0.9

a. Testing time too limited (30 min vs. recommended 40-60 min), data not representative of all students
b. From Eggen et al. (2017)
c. Doran (1980), Kline (1986)

Difficulty index *P* describes the item difficulty as a proportion of students who answer a question correctly. Thus, the higher the value of *P*, the easier the question is for the population tested. When comparing the values for different groups, CCI 3.0 appears somewhat difficult (low difficulty index) when used as a pre-test in both NTNU and JYU, but becomes easier when used as a post-test. However, making the test easier to correct for pre-test difficulty might lead to ceiling effects when using the same test as a post-test (Persson 2015). Thus, a slightly "too difficult" pre-test may be more optimal, if the test is still discriminating and reliable enough to be used as pre-test. It is also worth noticing that for NTNU 2015 post-test group, assumed to consist of high-achieving students (Eggen et al. 2017), the difficulty index is already quite high. This indicates that if the test is to be used for different student groups as both pre- and post-test, the range of difficulty indexes from these five tests is appropriate.

Some of the individual questions do not fit well in the recommended limits for the statistical test values. For example, the item discrimination index *D* has some small and negative values that indicate that this particular question cannot discriminate well between students in the high and low scoring groups, i.e. most of the students are probably guessing the answer. The same problem for some individual questions is seen in the point biserial coefficient r_{pbc}, where a small or negative value indicates that there is very weak or even negative correlation between a student answering this particular question correctly and the same student doing well in the test overall. However, most of the lower values are below the recommended limit only slightly, and there are only few individual values that are more drastically off limits. Except for the NTNU 2016 pre-test, questions failing all three statistical tests are rare, only two in the JYU pre-test, one in the JYU post-test, and one in the NTNU 2016 post-test. A detailed analysis of these questions is out of the scope of this conference proceedings, but a tentative evaluation of the problematic questions in JYU suggests that this might be due to the questions including symbols or terminology that are unfamiliar to the students at the time of testing.

In an overall view, according to the results from statistical analysis, CCI 3.0 can be considered a reliable and discriminating tool in the present contexts. There are some slight problems with a couple of individual questions that need a more detailed investigation. The test as a whole can, however, be used as a tool for evaluating first year chemistry students' chemistry concepts. As the test is statistically valid both as pre- and post-test, it could also be used to track the students' progress during the first semester, to study for example the effect of changing teaching methods or learning environments on students' chemistry concepts.

4. Conclusions

The Chemistry concept inventory CCI 3.0 developed at NTNU has been administered at two different Nordic universities as both pre- and post-test, and evaluated using statistical tests. Statistical analysis of the Chemistry concept inventory as a whole shows that its reliability and discriminatory power are at an appropriate level. The difficulty level of the CCI is higher when used as a pre-test when compared to post-test, but is acceptable for both purposes. A few individual questions need a more detailed analysis concerning their applicability for individual item analysis. The test as a whole can however be applied within the context of general chemistry courses at NTNU, JYU, and possibly other Scandinavian universities.

The inventory aims at functioning as a tool for adjusting teaching practices in general chemistry courses. Results of the tests could also be used to identify strengths and weaknesses in the students' understanding, thus helping students to focus on the areas in need for improvement. The present form of the chemistry concept inventory is shown to be a useful tool and will serve as a template for future versions as well as an inventory suitable for longitudinal studies. A more detailed analysis of the current version, CCI 3.0, using more test data, is in progress. Also, a future analysis of the answer distributions in individual questions will give us more data on student understanding of concepts and phenomena in general chemistry. The results will be used to develop the Chemistry concept inventory even further.

The CCI 3.0 questions can be accessed on request to the authors. Please contact Tiina Kiviniemi in JYU for questions in Finnish and Per-Odd Eggen in NTNU for questions in Norwegian.

References

Adams, W. K., Wieman, C. E., Perkins, K. K., & Barbera, J. (2008). Modifying and Validating the Colorado Learning Attitudes about Science Survey for Use in Chemistry. *Journal of Chemical Education*, *85*(10), 1435–1439.

Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the State of the Art for Measurement in Chemistry Education Research: Examining the Psychometric Evidence. *Journal of Chemical Education*, *90*(5), 536–545.

Bowen, C., & Bunce, D. M. (1997). Testing for Conceptual Understanding in General Chemistry1. *The Chemical Educator*, *2*(2), 1–17.

Doran, R. L. (1980). *Basic Measurement and Evaluation of Science Instruction.* Washington, DC: National Science Teachers Association.

Eggen, P.-O., 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.

Gabel, D. L., & Bunce, D. M. (1994). Research on problem solving: Chemistry. *Handbook of research on science teaching and learning*, *11*, 301–326.

Hestenes, D., Wells, M., & Swackhamer, G. (1992) Force concept inventory. *The Physics Teacher, 30*(3), 141–158.
Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. Methuen.
Krajcik, J. S. (1991). Developing students' understanding of Chemical concepts. In Glynn, S. M., Yeany, R. H., & Britton, B. K. (Eds.), *The Psychology of Learning Science* (pp. 117–147). Lawrence Erlbaum.
Krause S., Birk J., Bauer R., Jenkins B., & Pavelich M. J. (2004). Development, testing, and application of a chemistry concept inventory. *ASEE/IEEE 34th Annual Frontiers in Education Conference.* Savannah*.*
Kuder, G. F., & Richardson, M. J. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160.
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.
Nakhleh, M. B. (1992). Why some students don't learn chemistry: Chemical misconceptions. *Journal of Chemical Education*, *69*(3), 191–196.
Persson, J. R. (2015). Evaluating the Force Concept Inventory for different student groups at the Norwegian University of Science and Technology. *arXiv preprint arXiv:1504.06099*.
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), 115–163.
Stavy, R. (1991). Using analogy to overcome misconceptions about conservation of matter. *Journal of Research in Science Teaching*, *28*(4), 305–313.
Stavy, R. (1995). *Learning Science in the Schools: Research Informing Practice.* (pp. 131–154). Mahwah, N.J:Lawrence Erlbaum.
Taber, K. S. (2009). *Progressing Science Education: Constructing the scientific research programme into the contingent nature of learning science*. Dordrecht: Springer.
Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. In Gabel, D. (Ed.) *Handbook of research on science teaching and learning* (pp. 177– 210). New York: Macmillan.