Concept Maps: Theory, Methodology, Technology
Proc. of the Fifth Int. Conference on Concept Mapping
Valletta, Malta 2012

CONVERGENT VALIDITY: CONCEPT MAPS AND COMPETENCE TEST FOR STUDENTS' DIAGNOSIS IN PHYSICS

Siv Ling Ley, Heiko Krabbe & Hans E. Fischer, University of Duisburg-Essen, Germany
Email: siv.ley@uni-due.de, heiko.krabbe@uni-due.de, hans.fischer@uni-due.de
www.uni-due.de/didaktik_der_physik/agfischer.shtml

Abstract. In almost every lesson teachers diagnose their students concerning knowledge and activities, but most of them do so without appropriate diagnostic instruments. With the use of concept maps, the present study aims to offer physics teachers a suitable diagnostic instrument with an adequate analysis form for an instant use in their lessons. In a preliminary step, we first pinpointed what kind of knowledge concept maps measure. Therefore a sample of 79 ninth-grade German secondary school students drew concept maps under the instruction of two different task formats and afterwards filled out a competence test. For both task formats the concept maps were evaluated with the same scoring system and were tested for convergent validation against the competence test on the physics topic of energy. The results show a positive correlation between these two instruments and figure differences between the different task formats.

1 Introduction

Diagnosing students is among one of teachers' basic responsibilities. It is needed to evaluate and to reflect students' activities and to use this information to further adapt instruction. However, conducting diagnostics, which formally means "the systematic collection and preparation of information with the aim to obtain knowledge about people" (Jäger, 2009, translated into English), and performing the diagnosis require a suitable diagnostic instrument. It should be highlighted here that physics teachers in Germany are in a particularly disadvantageous position because appropriate and standardized diagnostic instruments for their instruction have not yet been developed. Given this general dearth of diagnostic tools, it was thought that concept maps drawn by students might be an appropriate way to diagnose their knowledge structures. Concept maps do not need any preparation and at a first glance it seems that they are easy to assess. A review of research about concept maps and concept mapping shows that they are an often-used topic in the international scientific community. Ruiz-Primo and Shavelson (1996) summed up and illustrated a multifaceted picture of how and in which way concept maps are used in research. They concluded that concept maps can be classified by three categories: task format, response format, and scoring system. They exposed the huge variability each of these categories can have. In the same manner as tests with different answer formats (e.g. multiple-choice, short-answers, etc.) and scoring systems measuring different aspects of knowledge, these varied concept mapping methods can generate distinct representations of knowledge. Therefore, it is not remarkable that results in research concerning reliability and validity are not consistent (for an example, see McClure, 1999). Reviewing German research on concept mapping, Peuckert and Fischler (2000) also pointed out that concept maps with specific forms of task format, response format, and scoring system measure characteristics of knowledge that cannot be perceived with other methods, like tests. It is demonstrably common that concept maps can measure knowledge as declarative knowledge, but in which way concept maps may also provide access to other kinds of knowledge is still an open question. Results on convergent validity of concept maps compared to other methods are, therefore, considered to depend specifically on task formats and scoring systems (Schecker and Klieme 2000, 47). According to Schecker and Klieme (2000), concept mapping tasks on a more theoretical and abstract level seem to be more closely related to competence test results than to tasks on a level of concrete daily experience. The latter allows more insight into the question of how concepts are connected to everyday language and life (Sumfleth and Tiemann, 2000).

The aim of this study is to develop a specific concept mapping task format and scoring system that is suitable for ad hoc assessments in school lessons and allows individual diagnostics for students' competences. In order to make this format and system suitable, a time-saving but sufficiently reliable scoring system is needed. For this study, a concept map-evaluation sheet has been developed, allowing for a thoroughly holistic evaluation of concept maps. The holistic rating of the concept maps is cross-checked with semantic measures of the software AKOVIA (Pirnay-Dummer and Ifenthaler, 2010).

With this background, the following research question leads the present study:

What relations exist between task format, scoring system of concept maps and students' competences measured by a competence test?

It is assumed that a concept map scoring – done with a concept map-evaluation sheet – will correlate positively with a competence test. Secondly, it is expected that a task format on a conceptual level will correlate more with the competence test than a task format with an everyday context[1].

1.1 Exursus: Basis Concept "Energy"

A short excursus is necessary at this point to understand the notion of the basis concept of "energy" within the German education system.

The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany decided years ago to develop so-called basis concepts for the natural sciences in order to provide overall links between the different science subjects and their special topics. For example, the basis concept "energy" can be taught in different grades with different levels of abstraction and related to different subtopics of physics (like mechanics or electricity). While knowing different energy forms is the main point in grades 5 and 6, learning the concept of energy conservation is the focus in grades 10 and 11. It is assumed that learning "energy" is done in four steps – so-called competence steps (Liu & McKeough, 2005): 1) energy form and energy sources, 2) energy transfer and energy conversion, 3) degradation of energy, and 4) energy conversation. The longer the students stay at school and learn physics, the more differentiated their concept of "energy" becomes.

2 Method

2.1 Instruments

2.1.1 Competence Test

The competence test for measuring students' competences in "energy" has been based on the work of Viering (2012) and Neumann, Viering, Boone & Fischer (under review). Twenty-two multiple-choice single-select items from a Rasch[2]-scaled item pool from Viering's work were selected for this current study. In the items the students have to apply theoretical knowledge about energy in concrete daily situations. The selection of the items was adjusted for the ninth grade to the four competence steps as mentioned in the excursus.

2.1.2 Concept Map Task Format

Two different concept map task formats were used in this study. In task format A, students had to choose at least 10 words out of 25, which are appropriately to the ninth grade. Furthermore, they could use the rest of the 25-word list or could add completely new words. The 25-word list consists of words belonging to the concept of energy, (e.g. energy form, energy storage, kinetic energy, power, energy loss, etc.). Thus, task format A is on an abstract, conceptual level.

In task format B, students were first presented three photos showing three different everyday physics situations: a girl jumping on a trampoline, a cooking pot with water, and an electric torch. They were asked to draw a concept map about the concept "energy" using self-chosen words. After drawing this concept map, they received the same 25-word list as in task format A and were asked to add at least 10 words using a pen in a different color. In this way, task format B starts on a concrete everyday level and switches to the conceptual level afterwards. The use of different pens allows for investigations on the connection between these levels. It also gives insight into students' active knowledge before hints from the 25-word list are given. In some ways, this can be compared to testing the Zone of Proximal Development (Vygotsky, 1978).
The students had 35 minutes for each task format. The association phase in task format B lasted about 14 minutes.

2.1.3 Concept Map-Evaluation Sheet

As previously mentioned, a so-called concept map-evaluation sheet was designed to evaluate the concept maps for the two task formats. The evaluation sheet consists of 18 statements concerning the physics basis concept

[1] The task formats will be explained further.
[2] The explanation about Rasch analysis is not the focus of this paper. For further general information about Rasch, see Bond and Fox (2007).

“energy”, which have to be rated by a four-step Likert-scale. The following example shows the main structure of the sheet.

Dear Teacher, In the following, you will find statements which you need to evaluate based on the concept map your student has drawn. Please mark an X at every statement that applies to the student’s concept map. The student…				
	Strongly agree (3 pts.)	Somewhat agree (2 pts.)	Somewhat disagree (1 pt.)	Disagree (0 pts.)
1. …knows that objects can possess energy, e.g. a battery has electric energy.	☐	☐	☐	☐
11. …has identified that energy forms can be converted into each other, e. g. potential energy will be converted into kinetic energy or thermal energy will be converted into internal energy.	☐	☐	☐	☐
16. …has identified that degradation of energy proceeds only in one direction (irreversibility), e.g. thermal energy which is released into the environment cannot be used anymore.	☐	☐	☐	☐
17. …has identified that total energy of a (closed) system rests conserved and is balanced, e.g. energy loss does not really exist, because energy conservation weighs and energy cannot be produced and discreated.	☐	☐	☐	☐

Figure 1. Instruction for teachers how to use the concept map-evaluation sheet and example-items of the concept map-evaluation sheet.

2.1.4 AKOVIA and modal maps

Additionally, a fully automated evaluation of the undirected relations (neglecting their labels) uses the software AKOVIA (Pirnay-Dummer and Ifenthaler, 2010). This software generates an average modal map of all individual maps (task format A and B) and compares each map to it. The creation of a modal map is after Hucke and Fischer (2000) a help to make sure that all linkages of the individual maps are used in an equal way. We gave the software a so-called 10%-limit which means that only propositions which can be found in at least 10% (more than 7 maps) of all 79 maps become part of this modal map. This fairly low limit had to be chosen due to task format B, which forced students to create their own vertices. This causes less overlap between the maps. Although the two task formats generate rather different maps, we wanted to generate a single modal map for both, in order to be able to compare them. The 10% limit seemed to be appropriate for that purpose, although for task format A higher limits would have worked as well.

For this study only the semantic indicators which AKOVIA calculates are important. The conceptual matching of an individual map is computed as the sum of the vertices that are semantically similar to the modal map. It is a general indicator of an accurate understanding compared to the modal map. The propositional matching calculates the number of propositions (vertex-edge-vertex) compared to the modal map. Only the connections between the vertices are compared, ignoring the edge labels. Propositional matching indicates a correct and deeper understanding of the subject domain represented by the modal map (Ifenthaler, 2010). Thus the creation of the modal map is essential to both indicators.

2.2 Sample

Participants of this study were ninth-grade physics classes in two German secondary schools. Altogether N=79 students took part. It was assumed that most of the students would not know what concept maps are and would need practical experience in order to draw concept maps. Therefore, the students first received a 45-minute training with practical sessions about how to draw concept maps. A few days later, the students were asked to draw a concept map for the physics basis concept of "energy". Each class was divided into two groups for the two task formats – sample distribution for task format A N=40 students and for task format B N=39 students. After drawing the concept maps, the students filled out the competence test which also dealt with the concept of "energy".

2.3 Data preparation

Before going into detailed results, it must be mentioned what happened after the students drew the concept maps. Before starting the final calculations, all concept maps were rated by six students who were at the beginning of their advanced study period in physics education for lower secondary schools. Each of them evaluated all of the 79 concept maps by using for each concept map the previously presented concept map-evaluation sheet. The points for each item of the evaluation sheet were summed up to give the rating score. So, for each concept map six different scores from the six raters exist. For further analysis, the six scores have been averaged. The number of students' correct answers in the competence test was summed up, so that a general view concerning the basis concept of energy is done, not a differentiated into the four competence steps.
All calculations were done with the statistic program PASW 18, when not mentioned else.

3 Results

3.1 Concerning the interrater reliability in the concept map-evaluation sheet

One of the most challenging aspects of concept maps is achieving objectivity in the form of interrater-reliability. In the current study, six different raters evaluated all concept maps. The intraclass-correlation-coefficient (ICC) was chosen as a degree of reliability for interval scales. For the six raters an $ICC_{(M=6)}$ with .515 on a level of $p<.001$was calculated, which is acceptable.

3.2 Concerning effort in time

The six raters were asked about the time they needed to score the 79 concept maps. The time for scoring one concept map was in average 5.47 minutes ranging from 3.79 minutes to 8.35 minutes.

Rater	Time for score set of 79 concept maps (min)	Per-map scoring time (min)
1	405	5.12
2	300	3.79
3	360	4.55
4	660	8.35
5	420	5.31
6	450	5.69

Table 1: Time required for raters to score 79 concept maps.

3.3 Concerning convergent validity between concept map-evaluation sheet and competence test

In order to prove convergent validity, students' scoring in the competence test was correlated with their scores given in the concept map-evaluation sheet by the six raters. Table 2 presents the results both on a general level and differentiated in the two task formats.

Correlation between…

Competence Test and Concept Map-Evaluation Sheet	Competence Test and Concept Map-Evaluation Sheet in task format A	Competence Test and Concept Map-Evaluation Sheet in task format B
$r = .273^*$, $p< .05$	$r = .314^*$, $p< .05$	$r = .331^*$, $p< .05$

Table 2: Correlation on Pearson between the competence test and the concept map-evaluation sheet (averaged across all raters); significant level p<.05 marked by *; p<.01 marked by **.

There is a significant correlation between the two instruments – competence test and concept map evaluation sheet.

3.4 *Concerning convergent validity between relational analysis by AKOVIA and concept map-evaluation sheet*

By using the computer program AKOVIA an automated relational analysis of concept maps (Pirnay-Dummer & Ifenthaler, 2010) can be compared with the results of the scores given in the concept map-evaluation sheet. For this comparison only semantic measures are considered.

The table below presents the results of this comparison by using a correlation-calculation on Pearson.

			Task Format A (N=40)	***Task Format B (N=39)***
		Score on Concept Map Evaluation Sheet (averaged across all raters)	Score on Concept Map Evaluation Sheet (averaged across all raters)	Score on Concept Map Evaluation Sheet (averaged across all raters)
Semantic Indicators (AKOVIA)	Conceptual Matching	*r* = .484** *p*< .01	*r* = .688* *p*< .05	*r* = .194 *p*= .237
	Propositional Matching	*r* = .132 *p*= .245	*r* = .092 *p*= .572	*r* = .060 *p*= .719

Table 3: Correlation on Pearson between the relational analysis by AKOVIA and the scoring in the concept map evaluation-sheet; significant level p<.05 marked by *; p<.01 marked by **.

The indicator conceptual matching shows a good agreement with the evaluation sheet across all maps and especially for task format A, but not for task format B. There is no correlation between the propositional matching and the score on the evaluation sheet.

3.5 *Concerning convergent validity between relational analysis by AKOVIA and competence test*

Analog to the comparison previously, a correlation calculation for triangulation between AKOVIA's semantic indicators and the score on the competence test was done. The results in the table show the supposed results. The competence test cannot be described by the semantic indicators which is comprehensible.

			Task Format A (N=40)	***Task Format B (N=39)***
		Score on Competence Test	Score on Competence Test	Score on Competence Test
Semantic Indicators (AKOVIA)	Conceptual Matching	*r* = .112 *p*= .327	*r* = .178 *p*= .271	*r* = .067 *p*= .686
	Propositional Matching	*r* = -.037 *p*= .747	*r* = .098 *p*= .548	*r* = -.103 *p*= .429

Table 4: Correlation on Pearson between the relational analysis by AKOVIA and the scoring in the competence test; significant level p<.05 marked by *; p<.01 marked by **.

3.6 *Concerning task formats*

By using a t-test for independent samples, the analysis for the two task formats shows no significant difference between the two student groups on the competence test. Keeping in mind the theory about what concept maps measure, there is a significant difference between the two student groups and therefore the two task formats. On average, students who worked with task format B received significantly more points in their rating than the students who worked with task format A. The following table sums up the results:

		Task format A (N=40)	*Task format B (N=39)*
	Score on competence test	*M*=9,15, *SD*=0,59	*M*=8,62, *SD*=0,49
	t-test	*t* (77)=0,693, n.s., *p*=.49, *d*=0.33	
Averaged on all raters	Score on concept map-evaluation sheet	*M*=7,03, *SD*=0,61	*M*=9,59, *SD*=0,51
	t-test	*t* (77)=-3,20, *p*<.01, *d*=0.66	

Table 5: Students' scoring divided into the two different task formats and instruments; t-test for independent samples and Cohen's *d* for effect size.

4 Discussion and conclusion

A primary purpose of this study was to find out whether a specific task of concept mapping (task format A and task format B) with a specific scoring system (concept map-evaluation sheet) measures the same ability that a competence test does. The interrater-reliability between the six raters is acceptable, but from a scientific perspective, upgradable. It should, however, be highlighted that the concept map analysis done with the concept map-evaluation sheet is holistic, and high interrater-reliabilities are not necessarily to be expected.

To date, the results on convergent validity show a slight positive connection between the two instruments – competence test and concept maps. It can be assumed that concept maps in the form in which they were used in this study can represent competences in the basic concept "energy".

Compared to the automatically calculated semantic indicators of AKOVIA, the rating with the evaluation sheet worked better. The needed time is almost acceptable time. The conceptual matching strongly depends on the use of equal labels in the vertices. In task format A a list of possible labels is provided, whereas in task format B the students are allowed to choose the vertices on their own. Thus, there will be little overlap in the modal map for task format B, due to the fact that synonyms like 'kinetic energy' and 'energy of motion' might be used. For the same reason propositional matching does not work, since the vertices are free to choose and differ too much. Taking the conceptual matching as an indicator for accurate understanding the correlation of .688 to the score on the evaluation sheet in case of task format A means that holistic rating can estimate the understanding properly. Additionally, holistic rating can take into account propositional information, too.

Task format A correlated with the competence test roughly as much as task format B did, though the two task formats differ in general. While task format A is more abstract with the technical terms which had to be used and accounts more for structured knowledge, task format B is more applied and accesses more the contextual application of knowledge. Both seem to be important for the competence test in equal manner. Therefore the similar correlations are not contradictory to each other. As presented in the results, students who worked with task format B received more points for their concept maps than those who worked with task format A. This suggests that concerning the content task format B creates more meaningful maps than task format A.

For general work scenarios, the concept map-evaluation sheet needs to be shortened because an 18 item-long sheet is not feasible for school[3], and the interrater-reliability should be matched to the scientific quality criteria, e.g. teachers judge the existing sample of concept maps with the concept map-evaluation sheet. With the teachers' rating the interrater-reliability could grow.

But referring to didactical relevance (see also Helmke, 2009), it would be acceptable if teachers' diagnoses were somewhat less precise, provided that teachers are aware of this and can control it themselves. The didactical relevance of the diagnosis is more important for the teachers than the accuracy of the diagnosis is. Therefore the scoring of concept maps with evaluation sheets seems to be a useful method for practice in school.

5 Acknowledgement

We thank Dirk Ifenthaler from the University of Mannheim (Germany) for the creation of the modal map and the calculation of the semantic indicators with his software AKOVIA.

6 References

Bond, T. G., Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* Mahwah: Lawrence Erlbaum Associates, Inc., Publishers.

Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung de Unterrichts.* Seelze-Velber: Klett Kallmeyer.

Hucke, L. & Fischer, H. E. (2000). Wissenserwerb und Handlungsregulation im physikalischen Praktikum. In Fischler, H. & Peuckert, J. (Hrsg.): *Concept Mapping als Forschungsinstrument in den naturwissenschaftlichen Fachdidaktiken.* Berlin: Logos Verlag.

[3] An explorative factor analysis has given hints for reducing the concept map-evaluation sheet by invariant correlation. But this is not focus of this paper.

Ifenthaler, D. (2010). Scope of Graphical Indices in Educational Diagnostics. In D. Ifenthaler, P. Pirnay-Dummer, N. M. Seel (Eds.). *Computer-based diagnostics and systematic analysis of knowledge* (pp. 213-234). New York: Springer.

Jäger, R. S. (2009). Diagnostische Aufgaben und Kompetenzen von Lehrkräften. In: K.-H. Arnold, U. Sandfuchs, J. Wiechmann (Eds.). *Handbuch Unterricht.* Bad Heilbrunn: Verlag Julis Klinkhardt. pp. 471-476.

Liu, X., McKeough, A. (2005). Developmental growth in students‘ concept of energy: Analysis from selected items from the TIMSS database. *Journal of Research in Science Teaching, 45* (5), 493-517.

McClure, J. R., Sonak, B., Suen, H. K. (1999): Concept Map assessment of classroom learning: reliability, validity and logistical practicality. *Journal of Research in Science Teaching 36* (4), 475- 492.

Neumann, K., Viering, T., Boone, W. & Fischer, H. E. (under review). Towards a Learning Progression of Energy. *Journal of Research in Science Teaching.*

Peuckert, J., Fischler, H. (2000). Concept Maps als Diagnose- und Auswertungsinstrument in einer Studie zur Stabilität und Ausprägung von Schülervorstellungen. In: H. Fischler (Ed.). *Concept Mapping in fachdidaktischen Forschungsprojekten der Physik und Chemie* (pp. 91-116). Berlin: Logos Verlag.

Pirnay-Dummer, P., Ifenthaler, D. (2010). Automated knowledge visualization and assessment. In D. Ifenthaler, P. Pirnay-Dummer, N. M. Seel (Eds.). *Computer-based diagnostics and systematic analysis of knowledge* (pp. 77-115). New York: Springer.

Ruiz-Primo, M. A., Shavelson, R. J. (1996). Problems and Issues in the Use of Concept Maps in science assessment. *Journal of Research in Science Teaching 33* (6), 569- 600.

Schecker, H., Klieme, E. (2000). Erfassung physikalischer Kompetenz durch Concept-Mapping-Verfahren. In: H. Fischler (Ed.). *Concept Mapping in fachdidaktischen Forschungsprojekten der Physik und Chemie* (pp. 23-49). Berlin: Logos Verlag.

Sumfleth, E., Tiemann, R. (2000). Own Word Mapping-Ein alternativer Zugang zu Schülervorstellungen. In: H. Fischler (Ed.). *Concept Mapping in fachdidaktischen Forschungsprojekten der Physik und Chemie* (pp. 179-203). Berlin: Logos Verlag.

Viering, T. (2012). *Unpublished dissertation*. IPN Kiel at University of Kiel.

Vygotsky, L. (1978). Interaction between learning and development. Reprinted in: M. Gauvain, M. Cole (1997, Eds.). *Readings on the development of children* (pp. 31-36). New York: W. H. Freeman and Company.