University of Twente
Enschede - The Netherlands

Construction of a Structural Knowledge Test

Matthias Matzner
August 2008
Docents: Dr. Hannie Gijlers, Dr. Hans van der Meij
Faculty of Behavioral Sciences, University of Twente, the Netherlands

A general definition of structural knowledge has not yet been established. This article provides a working definition of structural knowledge (related to the concept of deep understanding) from literature analysis. Common knowledge tests are compared on the theoretical basis of this working definition. The following comparison is done via an altered version of the SWOT (strengths, weaknesses, opportunities and threats) evaluation paradigm. This comparison between the different methods is further analyzed to generate design rules that can be applied to create a test that accounts better for the working definition of structural knowledge. These rules are then used exemplarily to create a test on a school course of chemistry. The test is initially tested at a small scale to explore its possibilities and then evaluated by the same SWOT analysis that was used earlier. The findings are mostly of prospective value for the paper at hand due to time constrains, small sample size as well as lack of knowledge about the test's reliability and validity. Finally, future research and implications are discussed.

Introduction. This article will describe how to construct a structural knowledge test. Since there is no clear consent about how to define structural knowledge this article will handle its own working definition.

The National Research Council [NRC] (see Bransford, Brown, & Cocking, 2000) states that the deeper knowledge is conceptualized the better it can be used to solve a related problem and the easier it is to modify the knowledge to counter novel problems. Despite the possibility, current teaching methods do not focus on the development of deeper understanding. And if only deep understanding renders a chance to excel in the job situation and work creatively (Bereiter, 2002) then it is advisable to have tools that monitor progress of the development of deep understanding. In other words, deep understanding needs to be made assessable (Carver, 2006).

The first section of the research paper at hand defines structural knowledge as composed of different knowledge types and, at high levels of structural knowledge, constituting deep understanding. After this definition follows an analysis to determine how structural knowledge is assessed at the moment. The results from that analysis will be used to create guidelines which need to be followed when assessing structural knowledge. Then these guidelines will be used to create an example test.

Foreclosing some results, I have to state that an opportunity emerged during writing. Since structural knowledge is, within the boundaries of this research paper, understood as composed of several aspects, the resulting multidimensional score of the test taker's structural knowledge might aid the diagnosis of problems for learner and teacher alike. This paper might thus be a progression of theory into praxis, that learner and teacher can get a better basis for communicating and hence interact with better results.

The Theory behind Structural Knowledge. According to the sources stated in the previous section supporting the development of deep understanding is important. Deep understanding, which will be understood as high amounts of structural knowledge in the context of this research paper, seems to play a crucial role or may even be the goal in learning. But up to this point the term structural knowledge remains unclear. In this section we will try to answer the first sub-question: How can we define the term structural knowledge in more precise phrases, i.e. what distinguishes "deep" knowledge from the supposedly shallow knowledge? – In the context of this research paper structural knowledge will be understood as consisting of four aspects: Structural knowledge is 1) of a declarative, episodic and/or procedural quality, 2) structured by semantic structure which are also part of the structural knowledge, 3) encoded verbally and/or non-verbally, and 4) its own meta-structure. A multidimensional score that comes forth from this definition might aid the diagnosis of the source of problems for the learner and teacher alike. Since this definition is merely based on theory it can only serve as a working definition which might need to be refined (see "General Discussion"). The remainder of this section will shed some light on the underlying theory of this working definition.

According to Tulving (1972) memory can generally be divided into two parts, namely declarative and episodic memory. The latter concerns knowledge that is bound to historical and personal dates, the memory of the fall of the "Berlin Wall" or ones own 20th birthday are examples of episodic memory. The former part of memory concerns knowledge that declares the states and properties of things and ideas, an example of content knowledge could be the property that 1) the element Sodium is chemically abbreviated Na. A second fact is that 2) cooking Salt has the chemical formula NaCl. To make sense of two bits of information one has to know that these concepts are interrelated, the learner has to have knowledge of the structural connection between these concepts. Someone with shallow, unstructured knowledge might only be able to answer questions regarding the single facts as they were encoded, but cannot flexibly make interferences between these facts. When, for example, Sodium is mixed with an unknown substance and the product of the reaction isn't salty, then someone with correct understanding of the interconnection of the two facts should be able to link these into the response that the unknown substance was not Cl (i.e. Chloride). The structural component of this knowledge domain is at first that these two concepts belong together and at second how these concepts work together. The first argument might be viewed as a judgment of relevance of this knowledge domain and the second argument might be viewed as the quality of the interrelation of the concepts.

A study conducted by Willingham, Nissen and Bullemer (1989) adds a third type, procedural memory. They repeated an earlier study done by Nissen and Bullemer (1987) while systematically choosing different subjects. They conducted the first with people with normal memory capability and in the second study participants were patients with memory disorders. The task was the same: Participants were asked to perform a serial reaction time task. The findings show that additional training increased both procedural and declarative knowledge of the sequence. The results show that development of knowledge in one system is not too dependent on knowledge in the other system and thus it can be assumed that there has to be something such as a procedural memory. Procedural knowledge comes into play when a task needs to be performed or, in other words, something has to be manipulated. When, following the above example of chemistry, the product of a reaction has to be predicted, procedural knowledge about the calculation is needed.

Jonassen and Wang (1993) define structural knowledge as the knowledge of how concepts within a domain are interrelated. This definition is similar to the definition of Preece (1976) when he equates structural knowledge to cognitive structure of knowledge; “Structural knowledge is also known as cognitive structure, the pattern of relationships among concepts in memory”. So the structural knowledge in the example of chemistry does not only consist of the mere facts 1) Sodium is chemically abbreviated Na and 2) cooking Salt is NaCl), but also incorporates the inferences that these two facts pose in conjunction. In other words, the facts have to relate to each other in one or the other way. These relations (or links) may vary in strength and implication, and structure the knowledge with different semantic connections and are what we call, along with the knowledge itself, structural knowledge.

When an array of facts is seen in respect of a certain semantic connection type then this is called a semantic network. In order of his research on artificial intelligence John F. Sowa (1987) distinguished between different types of semantic networks. The distinction was made due to the structure that was given to the incorporated knowledge. He proposed, for instance, definitional networks that structure the knowledge by hierarchy, which implies heredity of the properties of every subtype to its supertypes.

Another semantic network type, according to Sowa (1987), is the implicational network in which beliefs, causality or interference is the properties that give structure to the knowledge. Nevertheless it seems questionable if structural knowledge in humans is strictly organized in a certain manner and therefore can be categorized into a certain type of semantic networks. Due to the assumption that the human mind is not homogeneously structured it can be deducted that the human knowledge can be classified as a hybrid semantic network. The hybrid networks that are also proposed by Sowa are a lot more feasible when it comes to human minds instead of artificial intelligence. Sowa’s results are applicable here by aiding the distinction between different types of interrelations. Returning to the example, again, the interrelation of fact 1) and fact 2) can be defined as an *executable* interconnection. This means that the function of one concept influences the function of the interconnected concept.

In contrast to Sowa’s semantic networks Rumelhart and Ortony (1977) hypothesized in their Schema Theory that any knowledge stored within memory is represented as elements and attributes which are interrelated and thereby constitute schemata. Schemata are comparable to semantic networks for their structure, but are not bound to any specific type of interconnection. They can best be seen from Sowa’s point of view as hybrid semantic networks. All schemata in total about a specific topic form a hybrid semantic network. Sadoski, Paivio and Goetz (1991) offer criticism to this theory with the Dual Coding Theory. The authors claim that research done over the Schema Theory is easier and more consistently explained with this theory then with the Schema Theory. The Dual Coding Theory suggests that “cognition consists of two separate but interconnected mental subsystems”, namely, a verbal and a non-verbal system. This insight enhances what we already know about semantic networks in general. Thus, it might be that the logic structure is given by the types of semantic networks as proposed by Sowa (1987), but there is a more detailed differentiation to be made on encoding specificity (either verbal or non-verbal) of the knowledge at hand while keeping in mind that there are recognizable schemata within these networks.

But structural knowledge is not complete without meta-structure. According to Boshuizen and Schmidt (1992) experts, who in general supposedly have high levels of structural knowledge, structure their knowledge into big compartments, a natural process that is called knowledge encapsulation. The meta-structure, the knowledge encapsulations, enhances the knowledge and thus also is part of structural knowledge. Conclusively structural knowledge also consists of the knowledge about its meta-structure.

Aggregated from these theories and studies the definition of structural knowledge that this paper will work with goes as follows: structural knowledge is 1) of a declarative, episodic and/or procedural quality (Tulving, 1972; Willigham, Nissen & Bullemer, 1989), 2) structured by semantic structure[1] which are also part of the structural knowledge (Preece, 1976; Jonassen & Wang, 1993) 3) encoded verbally and/or non-verbally (Sadoski, Paivio & Goetz, 1991), and 4) its own meta-structure (Boshuizen & Schmidt, 1992). The following section will focus on how structural knowledge is assessed by common methods.

Structural Knowledge Assessment. The introduction illustrated the need for methods to assess structural knowledge. This section will evaluate if structural knowledge can be tested well at the moment. Currently there are many different knowledge assessment methods available where some seem to have the potential to assess structural knowledge. A selection was made from tests that could measure for structural knowledge, namely, card-sorting tasks, HyperCard stacks, tree constructions, dimensional representations (cognitive maps), pathfinder nets, verbal tests or structured essay questions (Jonassen & Wang, 1993; de Jong & Ferguson-Hessler, 1986; Jonassen, Beissner & Yacci, 1986; Gijlers, 2005).

The assessment methods now will be explained shortly by their test design and its underlying theory (i.e. test construct). The authors corresponding to the analyzed tests can be found in Table 1. However, some tests have a large number of different versions (e.g. concept maps). So the analysis had to focus on presenting the core of the tests rather than on presenting their diversity.

With Concept Maps in general the test taker is asked to map his/her knowledge onto a sheet of paper or a similar medium. The test taker is required to express the relatedness of two or more concepts in graphic distance to one another. This procedure is supposed to reflect the organization of knowledge within the test taker (Vargaz and Alvarez, 1992). Concepts ought to be linked correctly and sometimes the distance between concepts is also scored as proximity data.

Tree Constructions requires the test taker to attribute a number according to the perceived semantic proximity (i.e. strength of the connection between) between two terms. This slowly builds a tree of connections between provided terms. The final structure of the tree resembles the cognitive hierarchy that the knowledge is ordered by.

With Pathfinder Nets two concepts are presented and the test taker is asked to rate their proximity. This continues until enough data was collected to algorithmically calculate a graphical representation of the proximity of the concepts. Proximity data reflects relative structure of the knowledge. This is then compared to the proximity data of experts.

Card-Sorting Tasks consist of cards on which concepts are printed. The task is to sort these cards into piles. The card piles reflect the knowledge structure of the test taker. The results are then compared to the card sorting of experts.

HyperCard stacks are comparable to Card-sorting Tasks as they build upon the same premises, but are digitalized. The test taker has to create so called cards that contain information about a concept and link them to other cards/concepts. The links of the cards mirror the structure of the knowledge within the test taker.

The What-If test items supposedly test for intuitive knowledge (Swaak and de Jong, 1996). In a What-If scenario the test taker intuitively recalls the relevant knowledge and applies it to solve the problem. First a story is presented and then the test taker has to make the correct prediction about the outcomes.

This section described the tests by their test design and its underlying theory. In

---

[1] Semantic structure can further be distinguished into several different types (Sowa, 1987) and may be comparable to hybrid semantic networks (Rumelhart & Ortony, 1977; Sowa 1987).

the following three sections the six tests will be evaluated on their fit on the four aspects of structural knowledge (see section "The Theory behind Structural Knowledge" for the argumentation behind the four aspects of structural knowledge). First, the evaluation paradigm will be described (see section "Alternation of the SWOT evaluation paradigm") followed by its specification into practical aspects on which the tests will be evaluated. The second section, the results of the analysis (i.e. "Table 1") will be discussed in the third section ("Discussion of Table 1").

Alternation of the SWOT evaluation paradigm. This section is aimed at explaining the analysis paradigm that was used to create Table 1. A strengths-weaknesses-opportunities-threats [SWOT] analysis is usually conducted to compare business options (Deshpande & Ashtikar, 2005) but it handles the elements more strictly than is needed and useful here. Since the theoretical framework that a structural knowledge test has to satisfy has been described earlier, the current structural knowledge tests will be compared at their capability of assessing structural knowledge within this framework and focuses primarily on the creation of new test design rules.

A working definition of structural knowledge was established in the section "The Theory behind Structural Knowledge". The research paper at hand poses that structural knowledge is 1) of a declarative, episodic and/or procedural quality (Tulving, 1972; Willigham, Nissen & Bullemer, 1989), 2) structured by semantic structure which are also part of the structural knowledge (Preece, 1976; Jonassen & Wang, 1993) 3) encoded verbally and/or non-verbally (Sadoski, Paivio & Goetz, 1991), and 4) its own meta-structure (Boshuizen & Schmidt, 1992). The four key aspects of structural knowledge will be rephrased into a list to fit the need for ease of assessment. The product of this rephrasing is twofold: on the one hand it now serves as an orientation aid for the comparison, while on the other hand it will later provide guidance through the construction of a structural knowledge test.

For a test that has to assess structural knowledge, as defined by this research paper, the following conditions have to be closely kept in mind; if these conditions are met, then they count as strengths and if they are not met, then they count as weaknesses.

1) The diverse types (declarative, episodic and procedural) of knowledge have to be tapped.
2) The semantic network of the knowledge has to be assessed by the links, the properties of the links and the structure that the links form.
3) Both visual as well as verbal material has to be part of the test.
4) The test has to offer the possibility to assess for meta-structure of the knowledge.

The external quality for the analysis is oriented on the following points and count as either opportunities or threats:

1) The time needed for creating/administering the test.
2) Ease of scoring (training, analysis needed?).
3) Discriminating value between shallow and deep knowledge.
4) The test needs to be reliable and valid.

Table 1. Analysis Table of current Knowledge Tests.

Analysis Table of current Knowledge Tests						
	Concept Maps	Tree Constructions	Pathfinder Nets	Card-Sorting Tasks	HyperCard stacks	Intuitive Knowledge Test
Strengths	The advantage is that the concept can be freely generated and have to be discovered by the test taker before making the concept.	An advantage of this method is the simplicity of the data that is accumulated. "Rapid method of obtaining proximity measures." ***	The structural organization of the knowledge is visualized as in tree construction but the computerized algorithm saves time that would else wise be needed to score the proximity data.	The Card-sorting Task Can identify concepts and terminology that are difficult to categorize.	This method offers a lot of possible configurations of all the items and is easy to alter to fit different purposes.	This test might tap intuitive knowledge that is hypothesized to be more evident for inquiry learning environments and is more closely associated with procedural knowledge than with declarative knowledge.
Weaknesses	Due to the free association of concepts the test taker might be able to visualize and reflect his own knowledge and in fact elaborate and improve his conceptual understanding of the topic.	This assessment method, however, does not clarify why certain terms should be included, while others should be excluded, as well as the amount of terms provided. Furthermore, the strength of the connections is ambiguous and therefore cannot be compared.	The proximity measure remains ambiguous and the structural organization of the knowledge thus has no clear meaning and needs to be inferred by the researcher.	The comparison of the card stacks with experts is arbitrary. The difference in card-sorting might come from a cohort effect.	The comparison of the HyperCard stacks with experts is arbitrary. The difference in HyperCard-sorting might come from a cohort effect.	Scoring is difficult to generalize and comparative worth of test scores is uncertain
Opportunities	Dimensional Representations may help in assessing the organization (e.g. structure) of knowledge. The discrimination value of cognitive maps is the greatest when the test taker is unsupported to complete the task.**	Tree Construction might be used to assess the hierarchical organization of knowledge and thus shed more light on the depth of knowledge.	If, after a test taker ordered two concepts, he/she would be prompted to explain the type of connection, then this method could reveal more about the knowledge structure.	Card sorting is easy to conduct and therefore carries low risk for the test-retest reliability.	The program can easily be altered to include any amount of concept and links.	Moderately easy to modify to fit other knowledge domains.
Threats	Concept maps can be a tool assess as well as to learn with. It has been stated that two-dimensional representation of knowledge might not be sufficient.	A disadvantage is that the problem-space of the tree construction is well-defined while the real life problem-space of what the tree has to be constructed from is not.	The programming of the algorithm takes a lot of expertise thus hindering the modification onto new knowledge domains.	Not sure that every expert sorts cards according to problem-type. Less face reliability, because it is a pen and paper test and thus less controlled for than a computerized test.	Nonetheless it allows no direct inference on the quality of the links between the concepts. The links have to be of a predetermined quality and thus render no results concerning the quality of these links.	The freedom of test administration may cause low test reliability. Testing effects are also likely to occur. Test validity is only partially supportive.
Authors	Tergan, Graber, & Neumann, (2006); *Vargaz & Alvarez (1992); Novak & Cañas (2006); **Ruiz-Primo, Schultz, Li & Shavelson (2000).	Jonassen, Beissner & Yacci (1986); ***Preece (1976).	Chen (1998); McGriff, (2001); Jonassen, Beissner & Yacci (1986).	de Jong & Ferguson-Hessler (1986); Monchi, Petrides, Petre, Worsley & Dagher (2001); Lucey, Burness, Costa, Gacinovic, Pilowsky, Ell, & Marks Kerwin (1997).	Jonassen & Wang (1993); Thioulouse, Dolédec, Chessel & Olivier (1994); Hannay (1992).	Swaak & de Jong (1996); Swaak & de Jong (2001) Hulshof (2001).

Discussion of Table 1. The previous section used the working definition of structural knowledge (see “The Theory behind Structural Knowledge” for details) to describe criteria that a test has to fulfill to be seen as measuring structural knowledge as seen in the context of this research paper. This section will discuss and draw conclusions from the juxtaposition of the working definition of structural knowledge and the knowledge tests that were included in this analysis. The first possibility that has to be analyzed is if one test suffices to measure structural knowledge alone. After that it will be discussed if a combination of some of the tests could measure structural knowledge (i.e. if a test could be made up from a combination of the tests or some components thereof). But before entering the discussion it has to be emphasized that even though a test might not fulfill the requirements to test structural knowledge, it still did not fail. The tests that were analyzed in Table 1 do not build specifically upon and thus are unlikely to match the working definition that this paper handles regarding structural knowledge. Also, the possibility exists that the scope of the literature review was not broad enough, but for as far as this research paper goes these results mean that there might be no test that measures for structural knowledge as defined by the working definition.

Regarding the test design it seems that in most cases the tests used for this analysis require familiarization with the test task. This means that the tests require the test taker to accommodate to the new environment in which the knowledge is supposed to be applied. But it has to be noted that the familiarization process to any form of new learning and/or testing material claims time (Löhner, Joolingen & Savelsbergh, 2003). This delay means that when a testing environment is too different from the learning environment then the test takers have to familiarize to the new environment and will likely not be able to perform as they could. However, the question remains if all tests that were created according to a specific test design do not minimize the need for familiarization. The point of critique is that the test design does not address this error source, initially. It is, for instance, possible within the What-if format to formulate very closely to the previously provided learning experience, and thus minimize the need for familiarization, whereas it might be more difficult to do the same with concept maps.

All theories underlying those tests aim at an indicator that measure one single dimension. The accumulated scores as indicators make it nearly impossible to define the different aspects of structural knowledge in their specific quality. When the working definition is compared to the tests then it becomes apparent that the tests that were included in this analysis are able to assess nearly half of the key aspects of structural knowledge. Since the current tests build upon one-dimensionality they are unable to define the rich spectrum of structural knowledge. This gains importance when the test is not simply supposed to give the test administrator a score to distinguish between the quality of the structural knowledge of test takers, but when the learner also should benefit by assessing his own problem source with the test and learning in general. One example would be that the learner encoded the meta-structure as declarative information and the one-dimensional test, thus, does not reflect real meta-structure. Supposedly a test that builds upon the four key aspects of structural knowledge can identify not only deep understanding, but also the predominance of a certain knowledge aspect and thereby a possible problem source.

Two conclusions can be drawn from the analysis. The first conclusion that there is currently no knowledge test that measures structural knowledge alone as it is understood in the context of this research paper. The second conclusion is that the tests, excluding the What-If, neither call upon a single aspect of structural knowledge. The intuitive measure of the What-If format seems promising regarding procedural knowledge as more implicit knowledge that is only accessible when correctly cued. The indicators of the other tests always pose an already combined measure of the different aspects of structural knowledge and are thus ineligible for the testing of structural knowledge.

This, however, does not mean that all of these tests are unfit to aid in measuring the rich spectrum of structural knowledge. Concept maps are a widely used method to visualize knowledge and can be used to map the knowledge domain that is being tested. One should expect low reliability if the test taker learned with a strategy that is similar to the assessment method. The complexity of structural knowledge itself and the prerequisites of proper testing render a difficult task. The safest method to assess structural knowledge thus would be to use a test that cannot be used as a strategy for learning. This can be seen as a criterion for quality, although the negative effect of not-fulfillment of this criterion is limited if these reactive arrangements are otherwise compensated for. The following section will try to create general construction rules to be able to measure for structural knowledge.

Premises of a Structural Knowledge Test. This section will focus on discussing the analysis and the relevant aspects that come forth from it for the creation of a structural knowledge test. The aspects of structural knowledge (see "The Theory behind Structural Knowledge" for details) will be dealt with consecutively to come to an iterative description of the creation of a structural knowledge test. The examples formulated in this section are of a very general nature for illustrative purposes only and are thus unrelated to the following section "Example Test Creation: Chemistry".

Since the paper at hand pursues a multidimensional approach to assessing structural knowledge, multiple scales are created for the aforementioned aspects of structural knowledge. These multiple scales will be created using the item creation guidelines of DeVellis (1991). The guidelines summarize test-theoretical elements into an intelligible check-off list. This facet might be more interesting when future research has to formulate instructions for easy reuse and potential practitioners. The most important points of the check-list are presented here:

1. Predefine precisely what you want to measure. Utilize considerable amounts of theory for support. Keep items as specific as possible.
2. Avoid particular long items. They hardly do any good.
3. Write in the style appropriate for the test takers.
4. Pay special attention to "double-barreled" items that contain more than one concept. These are only allowed when the multifaceted elements add to their respective scales.
5. Mix the (positive and negative) wording of the items so that it is sometimes correct to agree and disagree with the question (i.e. item statement).

The first aspect states that "Structural knowledge is composed of declarative, episodic and/or procedural knowledge". Declarative knowledge can be tested by asking about the term for a concept related to the knowledge domain that the test shall be about or vice versa. The knowledge domain may be trees and a declarative question that could be asked would be: "What is the chemical abbreviation for Fluorine?" The reverse of this question would be to state the name of the chemical abbreviation, Fl, and ask what chemical this abbreviation stands for.

Episodic knowledge can be tested by asking to complete a sentence from previously provided learning experience. Although it might be wise to limit the sources for this knowledge type to the ones that are frequently used and those that are permanently available, such as books or other learning material. A question might be of a fill-in-the-blank nature: „Ions are created at a chemical reaction. Metals and nonmetals react to an ____." (answer: ion compound) The learning material could have been a chapter of a book on chemistry that deals with the formation of ions.

Procedural knowledge can be tested by posing a problem stemming from the previously provided learning experiences, which includes some kind of action rather than explanation or allocation of the correct answer to solve the problem. The problem might be to determine the outcome of a chemical reaction of Sodium and Chloride. To tap procedural knowledge all declarative information is given (e.g. Sodium is Na and is positively charged once, Chlorine is Cl and negatively charged once) and the test taker "only" has to have the knowledge how to process this information (metals and non-metals react to a salt, "metal non-metal" with adding the suffix "ide" and the electrical charges have to be balanced) into the correct answer (e.g. Sodium chloride).

The semantic structure of the knowledge is the second aspect of structural knowledge. Since the interconnections, as the name suggests, interconnect information, a combination of two questions into one can tap the existence of such an interconnection. Respecting that the knowledge domain permits it these two questions can be based upon any type of knowledge. This means that a question about declarative knowledge can be paired with a question about episodic, or procedural or another declarative knowledge. Since the interconnections of the knowledge also have to be distinguished into several different types it is advisable to keep track of the questions with a concept matrix (see Figure 1 for an example of a concept matrix). To prevent misinterpretation into a different semantic connection the test has to be specific on the premises, the relations and the circumstances of the questions. In other words, any question depends on the correct interpretation to enable the test taker to give the corresponding correct answer. The interpretation of the instructional text that a person has cannot be predicted. Thus, support is needed to guide the test taker to the correct interpretations of the questions, on one hand. Too much support, on the other hand, would slander the discriminating value between shallow and deep structural knowledge.

Regarding the third aspect, structural knowledge can be encoded verbally and/or non-verbally, it has to be assured that both visual as well as verbal material is part of the test. For the beginning items this material should stem from the previously provided learning experience. Feedback does not only mean the feedback that is given by the test instructor or the test program as a programmed response, but also the feedback that arises from cross-referencing between question texts and/or answers. Butler, Karpicke, and Roediger III (2007) studied the effect of feedback on learning. Their results let one conclude that feedback interferes with the assessment and should thus be avoided. This means, concerning the construction of a structural knowledge test, that the test design may not permit the test taker to answer questions by looking up information (verbal or non-verbal) from the text of another question or the answer.

To test for the last aspect of structural knowledge the test has to offer the possibility to assess for meta-structure of the knowledge. This might be done by posing questions about applying the principles of one knowledge domain to another knowledge domain. A distinction can be made between near and far meta-structure, depending on the probability that the domain is familiar or rather unfamiliar.

Method: Example Test Creation. The aim of this section is to clarify the use of the creation guidelines on an example knowledge domain. The example knowledge domain consists of several lessons about chemistry (Schroedel, pp 132-155, 2001). In general the test should start out easy, always keeping in mind to use instructions in self-referential style as Moreno and Mayer (2000) suggest and should refrain from giving feedback during the test (Butler, Karpicke, and Roediger III, 2007). The test can then gradually increase in difficulty until the last questions assess for meta-structure (see Appendix A for the "Example Test: Chemistry").

The first step in creating a structural knowledge test consists of mapping out the knowledge domain. For this the designer can use brainstorming methods and concept maps to clarify the relationship and possible names of the different concepts. The result of the brainstorming technique is then transferred to a knowledge matrix. This matrix serves the ease to create questions and the scoring system.

The knowledge domain of the chemistry course that the test is being designed for can be mapped out into formulas, elements, molecules, and the periodic table as the most basic concepts and their interconnections (see Figure 1: Basic Concept Map of the example Chemistry course). Alternative names for these concepts might include “table of the elements” for “periodic table” or “chemical compound” for “molecule”.

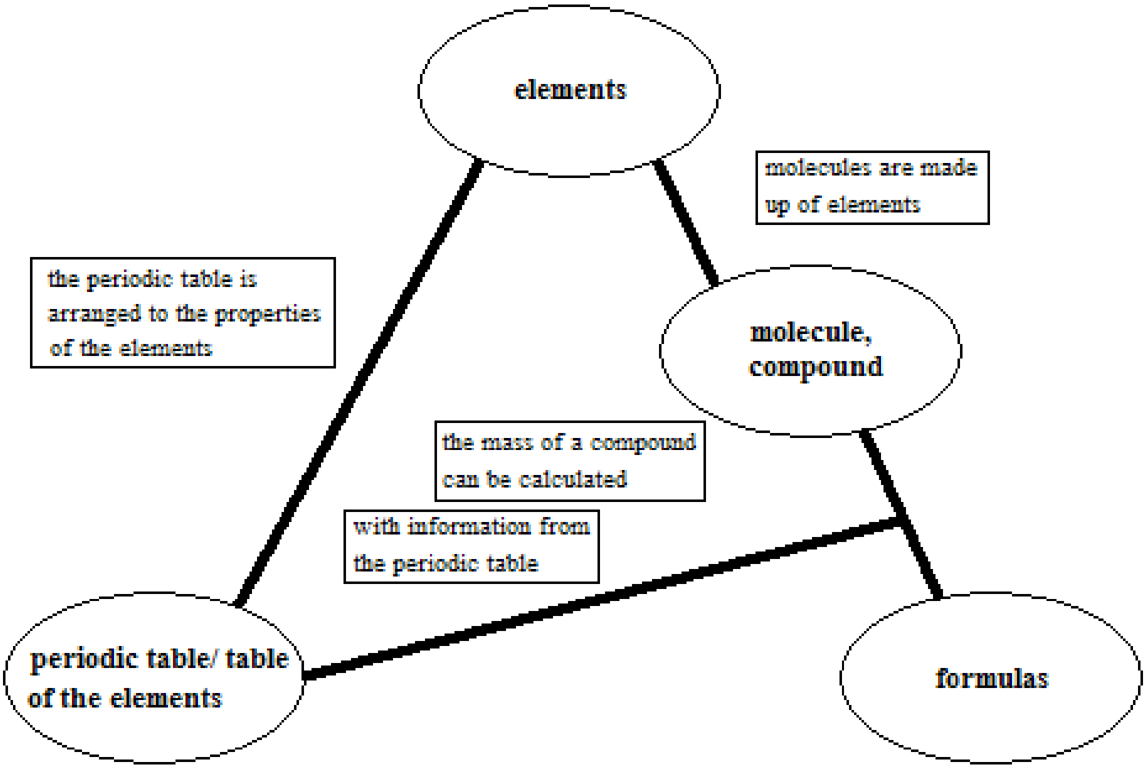


Figure 1: Basic Concept Map of the example Chemistry course

The concepts from the concept map along with the interconnections then have to be transferred into a concept matrix. The concepts are placed in the category spaces while the general interconnections fill the value spaces (Table 1: Concept Matrix of the example Chemistry course). Note that the space denoting the interconnections of the formulas and the elements (“X”) is left empty. That this connection is indirect does not mean that there is no link. The linkage is simply not definite enough (i.e. ambiguous) that one statement could represent the general relation of the concepts.

The assessment can begin to question the presence of knowledge about the content of the simple concepts. In the example for this paper, chemical abbreviations for elements, formulas and the structure of the periodic table consists of. Until research has clarified which question type is most suited for a certain aspect of knowledge, intuition should suffice to choose which combination to pick. It seems reasonable that the first questions should be yes/no or true/false questions that as well ask about technical terms as well as other declarative knowledge.

The next step consists of formulating questions, using the working definition (see “the Theory behind Structural Knowledge”) as a guideline. The first aspect of structural knowledge is that it is 1) of a declarative, episodic and/or procedural quality. The example test starts with yes/no-questions

Table 1: Concept Matrix of the example Chemistry course

	Elements	**molecule, compound**	**formulas**	**periodic table, table of the elements**
Elements	*	molecules are made up of elements	X (indirect link)	the periodic table is arranged to the properties of the elements
molecule, compound		*	mass (and such) of a compound can be calculated	to calculate the mass of a compound information from the periodic table is needed
Formulas			*	to calculate the mass of a compound information from the periodic table is needed
periodic table, table of the elements				*

about declarative knowledge if a certain chemical is abbreviated in a certain way (e.g. if „The chemical abbreviation for Bromine is Br", question # 3). Then it is tested if episodic knowledge about the periodic table is present by asking to indicate the location of a group (e.g. "Where are the alkaline earth metals in the periodic table?" question # 9). It is then asked to indicate if nonmetals and metals react together to salts and what suffix the compound gains when it is a salt, which is "ide". Procedural knowledge is difficult to tap in this domain since it is often entangled with declarative knowledge or is implicit. Question # 17 tries to tap this knowledge without necessitating the presence of another knowledge type. The content of the question is what to do, if a chemical reaction has to be formulated. This might tap the more implicit knowledge while keeping in check that possible lack of other knowledge would hinder completion of this question.

The second aspect of structural knowledge, the semantic structure, leads further to construct more difficult questions and is thus placed in later sections of the test. A combination of the knowledge about nonmetal-metal reaction and the knowledge about the properties underlying the periodic table has to be able to, if properly connected, predict the name of the product (question # 20). If knowledge about the nonmetal-metal reaction to a salt, which gains an "ide" suffix, is present and if that lithium is a metal and fluorine is a nonmetal, then a person who properly connects this information will be able to infer that the product of their reaction is called Lithium fluoride. Since structural knowledge is supposedly comparable to hybrid semantic networks this aspect has to be kept in check. The test designer has to assess whether the test taker has interpreted the question into the right semantic context. To realize this, the test designer has to incorporate answers that are only correct when set into another context. The amount of correct and incorrect answers that can be interpreted into another context has to be accumulated. In the example test answers that can be correct if interpreted in another semantic context are the first and the fifth answers of question # 17. The atom mass and the density of sodium and chlorine are not important, and thus incorrect in the context of the question, in a traditional chemical reaction. The first and the fifth answer are incorrect unless the context is interpreted otherwise to a task where the chemical reaction is also supposed to yield the mass of the products. The score of this subscale can serve as a measure for the misinterpretation, thus the name misinterpretation for this scale, of the test taker, as well as a measure for the ambiguity of the question in return.

The third aspect is that structural knowledge is encoded verbally and/or non-verbally. This aspect firstly leads there to that verbal and non-verbal material is used in the questions, while paying attention to

Table 2: Complete Scoring table of the example Structural Knowledge Test

Aspect of Structural Knowledge	Question	Maximum score
Declarative knowledge	1-8, 11 (x3), 12 (x3), 13-16, 18, 19, 21	24
Episodic knowledge	9 (x2), 10 (x2), 11 (x3), 12 (x3)	10
Procedural knowledge	14, 16, 17 (x4), 18-22	11
Linkage elements-periodic table	11 (x3), 12 (x3), 18, 19	8
Linkage formulas-molecule-periodic table	16, 17 (x4), 20-22	8
Misinterpretation	17 (x2)	-2
Verbal	1-8, 11 (x3), 17 (x4), 18-23	21
Non-Verbal	9 (x2), 10 (x2), 18-21	8
Meta-Structure	23	1
Total	1-23	91

keep the ratio of verbal and non-verbal material constant. When a previously provided learning experience did not have any verbal aspects in it, then it would be futile to use strictly verbal questions. See questions # 9 or # 18 for examples of how non-verbal information can be used to support the otherwise purely verbal character of the questions. These questions aim to tap episodic knowledge. While question # 9 stands for relatively free, i.e. unlinked, non-verbal knowledge, question # 18 resembles highly linked (namely with declarative knowledge) non-verbal knowledge. To generate a score from this aspect, every correct answer adds one point to either one or both scales depending upon their presence within the question.

Meta-structure, the fourth aspect of structural knowledge, can be assessed by asking if certain other domains (preferably ones that the test taker might be familiar with, for the near meta-structure, and other domains that he cannot be familiar with, for the far meta-structure) abide to the same scientific laws. The example test asks what ion-building has to do with chemical reactions (question #23). Since this domain was not dealt with in class before it has to be inferred and can provide a clue on if how far meta-structure is present.

The complete composition and therefore scoring of the test is revealed in Table 2. Each answer, if correct, adds one point to its according subscale. The only negative scale is misinterpretation where one negative point is added for an incorrect answer (i.e. answers #1 and #5 on question 17) due to misinterpretation.

This section gave an outlook on how the aspects of structural knowledge can be translated into test items. The following section will present the results of the test. The goals that are pursued by this test are completely explorative in nature. The mayor focus was to assess if the test format was comprehensible as such and if test takers grasp what the scores of the test (supposedly) say about their eventual deficits in learning.

Procedure. The testing took place in a Gesamtschule in Nordrheinwestfalen, a western state of Germany. The Gesamtschule is a school format teaching from ages 10-19 (5th to 13th grade) while differentiating according to performance between pupils. Only one participant did not want that his/her data was used while the data of four others had to be excluded due to statistical reasons, leaving 19 participants for analysis. The previously provided learning experience was defined by the book accompanying the course, (Schroedel, pp 132-155, 2001), and was then well discussed with the teacher to form the content of the test.

The participants had been given an information and consent form for the parents one week prior to the testing. The test session was held within one hour of class time (i.e. 45 minutes). The instructions, the testing, and the scoring of the tests all took 15 minutes, respectively. The instructions to the test included emphasis on the anonymity of the test results to ensure maximum participation and reducing tendency of socially desirable answers. Furthermore the instructions pointed out that the test results might aid the test takers in defining where their deficits lay (i.e. which aspect of structural knowledge is insufficiently developed). One participant mentioned a spelling mistake which indicates that there was no inhibition to report any mistakes or misunderstandings. Furthermore, there were no questions due to eventual misunderstanding of any questions which suggests that the questions were comprehensible overall. Then the scoring and the possible predictions about ones' problems with learning that could be made from the scores were explained in the remaining 15 minutes. Everyone attending that class received an unexpected small reward afterwards. The test content was made to fit the time limit of 45 minutes, which brings along limitations for the generalizability of the findings.

Results: Example Test. This section will deal with the results of the small scale experiment. The goals were to assess if the pupils understood the question format and the results of the test. Since the reliability and the construct validity are unknown the results cannot be generalized much further than to an explorative level. 19 (excluding five, see procedure) participants with mean age of 15,16 (SD 0,53) and a gender ratio of 11 (57,9%) males to 8 (42,1%) females voluntarily took part in the test during their normal class time. It was controlled for if gender or age had any influence on the results; but since this was not the case, the following analyses are presented without further regard to gender or age differences.

The participants mean grade for the chemistry course, as far as indicated, as well as the estimated overall grade average were statistically insignificantly correlated to the total reached score on the structural knowledge test ($r = 0.196$, $p = 0.451$; $r = 0.227$, $p = 0.337$, respectively). This could have several possible meanings: this could mean that the grade is not solely based on the level of understanding that the pupils acquired, e.g. the grade can be influenced by intangible factors such as behavior in the classroom; the previously provided learning experience could be different from the material that was used to create the test; the general formulation of the questions could be essentially different; the structural knowledge test could be unreliable, which cannot be investigated further within the boundaries of this research paper. At this point it has to be assumed that these points of concern play a minor role. Any speculations further into that direction would surely undermine the prospective intention of this research paper unnecessarily.

The means of the subscales (see Table 3: Example Test Means) show that no subscale has been either too hard or too easy. In other words, no subscale was redundant. Furthermore and more importantly, every subscale adds discriminating value to the test. Unfortunately the construct validity of the subscales as well as of the total scale (which could be seen as a measure for deep understanding, within the limitations of this research paper) is unknown. A reliability analysis was also omitted due to the small sample size. It would thus be futile to try to draw further conclusions from statistics concerning the subscales

Table 3: Example Test Means

Structural Knowledge Aspect of the Knowledge Domain	Maximum Score	Mean Score Acquired	Standard Deviation
Declarative knowledge	24	15.95	3.14
Episodic knowledge	10	5.53	2.3
Procedural knowledge	11	7.05	1.78
Linkage elements-periodic table	8	5.79	1.81
Linkage formulas-molecule-periodic table	8	3.89	1.66
Misinterpretation	-2	0.37	0.60
Verbal	21	12.84	2.22
Non-Verbal	8	4.79	1.437
Meta-Structure	1	0.68	0.478
Total	91	56.16	11.72

and the results, accordingly, have to be handled carefully.

The mean of the answers to the questions in the last section of the test, if the test results raised attention to their problems at learning and if the explanation raised attention to their problems at learning (see question 28 and 29 in Appendix A. Example Test: Chemistry), were between "moderately" and "good" (2,56 (SD = 0,984) and 2,59 (SD = 0,795) respectively, with 1 as "very good" and 5 as "not at all"). That the appraisal of the pupils was mainly positive seems promising regarding that a structural knowledge test might have mutual benefit for the teacher and the pupils alike. The explanation of the sub-scores raised many questions of how the problems of learning can be worked out. This seems to further support that the test format is interesting for the pupils themselves to learn about their problems in learning.

SWOT Analysis of the Structural Knowledge Test. To evaluate the structural knowledge test creation guidelines, the same SWOT analysis that the possible assessment methods for structural knowledge (see "Structural Knowledge Assessment") had undergone is conducted again.

The structural knowledge test measures (until empirically investigated, only theoretically) the diverse types (declarative, episodic and procedural) of knowledge and the other aspects of structural knowledge which can be seen as a strength of the example test. It might count as a weakness that the particular example test taps meta-structure and misconception by only one item each. Furthermore the linkage between molecule and element stays untapped, which was caused by lack of experience with the test creation process in general.

Opportunities for this kind of test seem to be that the time needed for administering the test is relatively low. The test can be completed in the time that any other common classroom test can. The pupil does not have to accommodate to a new test procedure or answer schema. The ease of scoring that results from the scoring table (see "Table 2: Complete Scoring table of the example Structural Knowledge Test" for an example) might aid in implementing such a test in a real education context.

A threat however, might be that detailed instructions are required that support a good test design and implementation. As the mistake with the lack of items for the linkage between molecule and element and the low amount of items of meta-structure and misinterpretation (compare "Table 1: Concept Matrix of the example Chemistry course" with "Table 2: Complete Scoring table of the example Structural Knowledge Test") has shown the quality of the test is highly dependant on the test designer. The possibly underspecified instructions might be the cause of this problem and should be formulated more strictly in future research. The scoring of the test is complicated until further specification of the instructions. Furthermore, the discriminating value between shallow and deep knowledge as well as the test's reliability and validity are unknown and have yet to be supported by empirical evidence.

Conclusions and Discussion. Structural knowledge seems to be a concept that might support refining the teaching and assessment methods. A problem seems to be that there is no commonly accepted definition for structural knowledge. The paper at hand tries to deal with this problem by implementing its own working definition.

In assessing structural knowledge, there is a lack of congruency; while some tests rely on indicators that measure some part of deep understanding, others build upon a single question type or task. The working definition that this paper uses regarding structural knowledge (i.e. deep understanding) postulates that there is more to understanding than mere meta-structure or mere applicability of the knowledge. Besides, this working definition has shown (at least on this explorative level) that it is comprehensible and creates interest.

The literature review started with sketching out how structural knowledge can be assessed. To compare these assessing methods the SWOT paradigm was chosen and then altered to fit the needs of this paper. The results of this literature review led to the formulation of guidelines for a successful test creation to measure for structural knowledge. These creation guidelines were then used to create a structural knowledge test to verify these creation guidelines and to examine whether a structural knowledge test is actually realizable. The results suggest that the test is realizable, even in a relatively short period of time, and might yield useful results by providing the teacher and the pupil with an informative basis.

The instructions may be crucial to ensure the maximum participation and to support the diagnostic character of the test. Guidelines might not be enough to fully guarantee an easy test creation process and fulfillment of the theoretical basis. The example test creation showed that it is easy, on the one hand, to create test items and to develop the scoring system, but it was rather difficult to keep track of every single aspect of structural knowledge. A step-by-step instruction could possible yield better results even with inexperienced test designers.

The pupils were eager to know where their deficits lay and confident that the test results and the explanation given to these results showed them where they lack understanding. This seems to support the assumption strongly that structural knowledge as well as the measure for it, a structural knowledge test, are comprehensible and aid in the diagnosis of the interrelation of learner and teacher.

In general, a structural knowledge test might be a feasible method to assess for deep understanding and to diagnose problems with learning. The lack of knowledge about the test's reliability and validity and the probably underspecified instructions, that are mere guidelines up to this point instead of rules, threaten the integrity and the possible benefits that arise from such a multidimensional measure for deep understanding as a structural knowledge test might be. Such a measure could refine the learning experience by shifting the focus from mere grades to the process of learning instead. The interest of the pupils in how to achieve a higher score in a certain subscale of the test can be seen as supporting this opportunity by raising the attention to the aspects of deep understanding (i.e. the single aspects of structural knowledge).

As stated many times, this research is explorative and has its limitations. The time constrains did not only have influence on the amount of knowledge that the test could assess. For a real carefully constructed structural knowledge test it might be advisable to have extended knowledge of the knowledge domain and the previously provided learning experience that the test refers to.

There are questions that remain unanswered. Although the data seems promising on this explorative level, the scale of the test might be too small to generalize any further than this. A larger sample with an extended structural knowledge test and a control group that is assessed by a normal classroom test might give better results since the grade for a course is not only composed of knowledge tests but also of other aspects of pupil performance. A larger sample could also provide evidence concerning the validity and/or necessity of the subscales and therefore the validity of the construct structural knowledge. To be repeatable over different knowledge domains the test design instructions have to be formulated by research and extended reliability analysis. From the quality of these instructions depends the speed and accuracy with which a new test can be made and thus if the test design is realizable in a real education context as such.

References

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Lawrence Erlbaum Associates.

Boshuizen, H.P.A., Schmidt, H.G., (1992). *On the Role of biomedical Knowledge in clinical Reasoning by Experts, Intermediates and Novices.* Cognitive Science: A Multidisciplinary Journal. Vol. 16, No. 2, pp 153-184, University of Limburg, The Netherlands.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds), (2000). *How People Learn: Brain, Mind, Experience, and School.* Washington, DC: National Academy Press.

Butler, A.C., Karpicke, J.D., and Roediger III, H.R., (2007). The Effect of Type and Timing of Feedback on Learning From Multiple-Choice Tests. Journal of Experimental Psychology: Applied. American Psychological Association, Vol. 13, No. 4, pp 273–281.

Carver, M.S., (2006). *Assessing for deep understanding*. In K. Sawyer (Eds.), Cambridge handbook of the learning sciences, pp. 205-224.

Chen, C., (1998). *Bridging the Gap: The Use of Pathfinder Networks in Visual Navigation.* Journal of Visual Languages and Computing, Vol. 9, pp 267 – 286.

Deshpande, S.G., & Ashtikar, R., (2005). *SWOT Analysis of Distance Education for testing its Suitability to impart technical and vocational Education.* ICDE International Conference, November 19 – 23, 2005, New Delhi. Government Polytechnic College, Nagpur, India.

Gijlers, H., & de Jong, T., (2005). *Confronting Ideas in Collaborative Scientific Discovery Learning.* Department of Instructional Technology, University of Twente.

Hannay, D.G., (1992). *Hypercard automata simulation: finite-state, pushdown and Turing machines*. ACM SIGCSE Bulletin, Vol. 24, No. 2, pp 55-58.

Hulshof, C.D., (2001). *Discovery of Ideas and Ideas about Discovery*. Thesis, Twente University Press, the Netherlands.

Jonassen, D. H., Beissner, K., & Yacci, M., (1993). *Structural Knowledge: Techniques for representing, conveying, and acquiring structural knowledge.* Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Jonassen, D.H., Wang, S., (1993). *Acquiring Structural Knowledge from Semantically Structured Hypertext.* Journal of Computer-Based Instruction.

Jong, T. De, & Ferguson-Hessler, M.G.M., (1986). *Cognitive Structures of Good and Poor Novice Problem Solvers in Physics*. Journal of Educational Psychology, Vol. 78, No. 4, pp 279-288.

Löhner, S., Joolingen, W.R. van, Savelsbergh, E.R., (2003). *The Effect of external Representation on constructing Computer Models of complex Phenomena*. Instructional Science, Vol. 31, pp 395-418.

McGriff, S., (2001). *Measuring Cognitive Structure: An Overview of Pathfinder Networks and Semantic Networks*. Conceptual Learning and Problem Solving, Pennsylvania State University, Educational Psychology 523.

Nissen, M.J., Bullemer, P., (1987). *Attention Requirements of Learning: Evidence from Performance Measures*. Cognitive Psychology, Vol. 19, No. 1, pp 1-32.

Novak, J.D., Cañas, A.J., (2006). *The Theory Underlying Concept Maps and How to Construct Them*. Technical Report IHMC CmapTools 2006-01.

Preece, P.F.W., (1976). *Mapping cognitive structure: a comparison of methods.* Journal of Educational Psychology, Vol. 68 pp.1-8.

Ruiz-Primo, M.A., Schultz, S.E., Li, M., & Shavelson, R.J., (2000). Comparison of the Reliability and Validity of Scores from two Concept-Mapping Techniques. Journal of Research in Science Teaching, Vol. 38, No. 2, pp. 260-278.

Rumelhart, D., & Ortony, A., (1977). *The representation of knowledge in memory*. In R.C. Anderson, R. J.

Sadoski, M., Paivio, A., Goetz, E.T., (1991). Commentary: A Critique of Schema Theory in Reading and a Dual Coding Alternative. *Reading Research Quarterly*, Vol. 26, No. 4, pp. 463-484.

Schroedel, (2001). *Chemie Heute S1 Gesamtband*, Bildungshaus Schulbuchverlage, pp 132-173.

Shavelson, R.J., (1972). Some Aspects of the Correspondence between content Structure and Cognitive Structure and Cognitive Structure in Physics Instruction. Journal of Educational Psychology, Vol. 63, No. 3, pp 225-234.

Sowa, J.F., (1987). Revised and extended version of an article originally written for the *Encyclopedia of Artificial Intelligence*, edited by Shapiro, S.C., & Wiley, (1987).

Swaak, J., & de Jong, T., (1996). Measuring Intuitive Knowledge in Science: The Development of the What-If Test. Studies in Educational Evaluation, Vol. 22, No. 4, pp 341-362.

Swaak, J., & de Jong, T., (2001). *Discovery simulations and the assessment of intuitive knowledge*. Journal of Computer Assisted Learning, Vol. 17, No. 3, pp 284–294.

Tergan, S., Graber, W., Neumann, A., (2006). *Mapping and Managing Knowledge and Information in Resource-Based Learning*. Innovations in Education & Teaching International, Routledge, Vol. 43, No. 4, pp 327-336.

Thioulouse, J., Dolédec, S., Chessel, D., & Olivier, J.M., (1994). *Ade Software: Multivariate Analysis and graphical Display of Environmental Data* in G. Guariso and A. Rizzoli, editors. Software per l'ambiente, Pàtron Editore, Bologna, pp 57-62.

Tulving, E., (1972). *Episodic and semantic memory*. Organization of Memory, Academic Press, New York.

Vargas, E.M., & Alvarez, H.J., (1992). *Mapping out Students' Abilities.* Science Scope, Vol. 15, No 6, pp 41-43.

Willingham, D.B. Nissen, M.J., Bullemer, P., (1989). *Title On the development of procedural knowledge.* Source Journal of Experimental Psychology: Learning, Memory, and Cognition. Vol. 15, No. 6, pp 1047-1060.

Appendix A. Example Test: Chemistry

Indicate if the following statements are either right or wrong

1. „The chemical abbreviation for Potassium is P.“
☐ Right
☐ Wrong

2. „The chemical abbreviation for Lithium is Lt.“
☐ Right
☐ Wrong

3. „The chemical abbreviation for Bromine is Br.“
☐ Right
☐ Wrong

4. „The chemical abbreviation for Helium is He.“
☐ Right
☐ Wrong

Answer the following questions by indicating the correct answer.

5. What is the chemical abbreviation of Magnesium?
☐ Mg
☐ Me
☐ M
☐ Mn

6. What is the chemical abbreviation of Sulphur?
☐ Se
☐ S
☐ Ss
☐ Sw

7. What is the chemical abbreviation of Argon?
☐ Arg
☐ Ar
☐ A
☐ G

8. What is the chemical abbreviation of Oxygen?
☐ S
☐ St
☐ O
☐ Arg

To answer the next questions indicate your answer through a line.

9. Where are the alkaline earth metals in the periodic table?

	I	II	III	IV	V	VI	VII	VIII
1								
2								
3								
4								
5								
6								
7								

10. Where are the noble gases in the periodic table?

	I	II	III	IV	V	VI	VII	VIII
1								
2								
3								
4								
5								
6								
7								

---

11. Which of the following elements belong to the halogen family? (multiple answers possible)
☐ Chlorine
☐ Sodium
☐ Hydrogen
☐ Helium
☐ Potassium
☐ Flourine
☐ Bromine

12. Which of the following elements belong to the alkaline metal family? (multiple answers possible)
☐ Chlorine
☐ Sodium
☐ Hydrogen
☐ Helium
☐ Potassium
☐ Flourine
☐ Bromine

To answer the next questions indicate your answer through crossing them on

13. What does it mean when an atom reaches noble gas configuration?
☐ the atom smells nice
☐ the atom is a gas now
☐ the most outer orbit is filled completely
☐ ions were created

14. What is the ‚product‘ of a chemical reaction?
☐ the starting reactants
☐ the result of the reaction
☐ the numbers on an atom
☐ the model

15. What is the difference between ion and atom?
☐ ions are negatively charged, atoms are neutral
☐ atoms are neutral, ions (+ or -) charged
☐ ions are green
☐ atoms are positively, ions are negatively charged

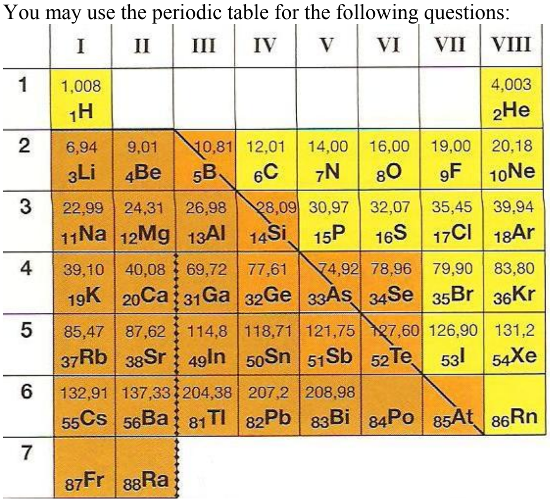
16.What is the general reaction schema of metals and non-metals?
☐ metal + non-metall ---> non-metall + metal
☐ metal + non-metall ---> salt (name + "ide") + rest
☐ metal + non-metall ---> non-metalloid + air
☐ 2Me + H2O ---> 2MeOH + H2

Please hand this page in when you are done.

17. What would you do if you wanted to formulate the chemical reaction of sodium and chlorine? (multiple answers possible)
☐ I look up what atom mass sodium and chlorine have
☐ I write sodium and chlorine on the left side of the reaction schema
☐ I look up what atom number sodium and chlorine have
☐ I look up how many electrons sodium and chlorine need to reach noble gas configuration
☐ I look up what density sodium and chlorine have
☐ I equate the amount of the starting reactants, according to how many electrons they need to reach noble gas configuration

---

You may use the periodic table for the following questions:



	I	II	III	IV	V	VI	VII	VIII
1	1,008 $_{1}$H							4,003 $_{2}$He
2	6,94 $_{3}$Li	9,01 $_{4}$Be	10,81 $_{5}$B	12,01 $_{6}$C	14,00 $_{7}$N	16,00 $_{8}$O	19,00 $_{9}$F	20,18 $_{10}$Ne
3	22,99 $_{11}$Na	24,31 $_{12}$Mg	26,98 $_{13}$Al	28,09 $_{14}$Si	30,97 $_{15}$P	32,07 $_{16}$S	35,45 $_{17}$Cl	39,94 $_{18}$Ar
4	39,10 $_{19}$K	40,08 $_{20}$Ca	69,72 $_{31}$Ga	77,61 $_{32}$Ge	74,92 $_{33}$As	78,96 $_{34}$Se	79,90 $_{35}$Br	83,80 $_{36}$Kr
5	85,47 $_{37}$Rb	87,62 $_{38}$Sr	114,8 $_{49}$In	118,71 $_{50}$Sn	121,75 $_{51}$Sb	127,60 $_{52}$Te	126,90 $_{53}$I	131,2 $_{54}$Xe
6	132,91 $_{55}$Cs	137,33 $_{56}$Ba	204,38 $_{81}$Tl	207,2 $_{82}$Pb	208,98 $_{83}$Bi	$_{84}$Po	$_{85}$At	$_{86}$Rn
7	$_{87}$Fr	$_{88}$Ra						

18. What kind of number is „1,008“ at oxygen?
☐ Atom mass
☐ Atomic number
☐ Ion binding behaviour
☐ Atom weight

19. What kind of number is „35“ at Bromine?
☐ Atom mass
☐ Atomic number
☐ Ion binding behaviour
☐ Atom weight

20. What compound is produced from lithium and fluorine?
☐ Lithium oxide
☐ Lithium fluoride
☐ Fluoscenting Lithium
☐ Fluor lithide

21. What compound is produced from magnesium and chlorine?
☐ $MgCl^2$
☐ $ClMg^3$
☐ Cl^2Mg
☐ Mg^2Cl

22. What product results from $H^2 + Br2$?
☐ 2HBr
☐ $2BrH^2$
☐ BrH
☐ Brom

23. How is ion formation related to chemical reactions?
☐ Ion formation makes chemical reactions generally possible
☐ Ion formation is important, but does only occur when a current is supplied
☐ not related, only atomic number and elemental group are important
☐ not related, ions are only formed when salt dissolves in water

---

Now I have some questions about you. As mentioned before, this is all anonymous and no other than me will see your answers.
Year of birth, months: ______________
gender: ♀ ♂
last (estimated) grade average: _._ _
Chemistry grade:

How would you evaluate own knowledge about the subject chemistry?
☐ very good ☐ good ☐ intermediate ☐ not so good ☐ not so good at all

How would you evaluate own skills about the subject chemistry?
☐ very good ☐ good ☐ intermediate ☐ not so good ☐ not so good at all

How well could you explain something from this chemistry course to someone else?
☐ very good ☐ good ☐ intermediate ☐ not so good ☐ not so good at all

How would you evaluate yourself as a pupil?
☐ very good ☐ good ☐ intermediate ☐ not so good ☐ not so good at all

- - - - - - - wait a moment - - - - - - -

Have the test results raised your awareness towards your problems with learning?
☐ very good ☐ good ☐ intermediate ☐ not so good ☐ not so good at all

Have the explanations raised your awareness towards your problems with learning?
☐ very good ☐ good ☐ intermediate ☐ not so good ☐ not so good at all

---

END of test. Thanks for your assistance