Validating domain ontologies: A methodology exemplified for concept maps

Christina M. Steiner* and Dietrich Albert

Abstract: Ontologies play an important role as knowledge domain representations in technology-enhanced learning and instruction. Represented in form of concept maps they are commonly used as teaching and learning material and have the potential to enhance positive educational outcomes. To ensure the effective use of an ontology representing a knowledge domain it needs to be validated. In this paper a previously presented validation methodology for concept maps is exemplified. Two different types of concept map validity are distinguished, referring to the correctness of the concept map’s content (content validity) and to the applicability of the concept map for its designated purpose (application validity), like its use in intelligent tutoring. To demonstrate the usefulness of the two validation types and approaches, they are illustrated by an empirical study. The content validity of a concept map on elementary geometry has been investigated by comparing it with empirically collected criterion maps through similarity measures. To demonstrate application validity an approach utilising methods of Knowledge Space Theory for predicting problem-solving behaviour has been applied. The obtained results show appropriate content validity as well as application validity for the given concept map and argue for the practical relevance of the proposed validation framework.

ABOUT THE AUTHORS
Christina M. Steiner completed her Diploma in Psychology and received her Dr.rer.nat at the University of Graz, Austria. She is a researcher at the Knowledge Technologies Institute of Graz University of Technology, Austria. Her research interests include knowledge and competence modelling, technology-enhanced learning, serious games, adaptation and personalisation, self-regulated learning, concept maps, and user-centred evaluation of information technologies.

Dietrich Albert has a postdoctoral degree in psychology and is the head of the Cognitive Science Section of the Knowledge Technologies Institute at Graz University of Technology, Austria. His main research areas are on modelling knowledge and competence, on developing comprehensive knowledge representation frameworks and ontologies, and on their application in technology-enhanced learning for personalising learning paths and for adaptively assessing knowledge and competencies of learners.

PUBLIC INTEREST STATEMENT
Ontologies describe the concepts of some area of interest or subject matter. Visualised in form of so-called concept maps, they are commonly used to represent knowledge in traditional education and in e-learning. Before using a concept map, it is essential to ensure that it correctly represents the respective knowledge domain and is suitable for its designated purpose. We demonstrate a methodology for validating concept maps by an example and study in the field of geometry. It has been investigated whether a given concept map on right triangles appropriately mirrors the concepts and interrelations of the subject matter (content validity). In addition, it has been analysed whether the concept map can be used to predict problem-solving performance on geometry exercises (application validity). The results of our study show that our concept map accurately represents the knowledge domain and intended application, and that the proposed methodology is suitable for practical use.
1. Introduction
Ontologies are knowledge representations providing an “explicit specification of a conceptualization” (Gruber, 1995, p. 908). A domain ontology identifies the key concepts, objects and entities that exist in some knowledge domain or area of interest and the relationships between them (Genesereth & Nilsson as cited in Gruber, 1995; Zouaq & Nkambou, 2008). Ontologies play a significant role for knowledge sharing and as knowledge models in instructional science, technology-enhanced learning, knowledge management and training (e.g. Gruber, 1995; Kickmeier-Rust & Albert, 2008; Zouaq & Nkambou, 2009). Approaches of ontology learning, i.e. the (semi-)automated extraction or generation of knowledge representations, aim at dealing with the bottleneck inherent in the manual creation of domain ontologies and with the difficulty of modelling the knowledge that is actually relevant for a specific domain (e.g. Granitzer et al., 2007; Hazman, El-Beltagy, & Rafea, 2011; Huang, Yang, & Lawrence, 2015; Lau, Dawei Song, Yuefeng Li, Cheung, & Jin-Xing Hao, 2009; Maedche & Staab, 2004). Domain ontologies in an educational context describe the knowledge domain that is the subject of learning and, in this connection, are usually so-called lightweight ontologies (Corcho, Fernández-López, & Gómez-Pérez, 2003; Lau et al., 2009). These domain ontologies are commonly used and depicted in form visual representations of concept or knowledge maps (e.g. Šimko & Bieliková, 2009). In this paper we focus on concept maps representing domain ontologies. The application of concept maps has generated a lot of attention and interest in educational researchers and instructional practitioners in the last decades. Concept maps have become a key instrument in traditional instruction and e-learning, they play a key role as teaching and learning material, and they have the potential to enhance positive educational outcomes (for an overview see e.g. Ifenthaler & Hanewald, 2014; Steiner, Albert, & Heller, 2007).

Concept maps are tools for capturing and presenting semantic knowledge and its conceptual organisation (for an overview see e.g. Novak, 1998; Steiner et al., 2007). They specify the concepts of a knowledge domain and the relationships among them and thus, provide a natural way of expressing and presenting domain ontologies. Mathematically defined, a concept map, is a directed graph consisting of a finite, non-empty set \( C \) of nodes, which represent the concepts of a knowledge domain, i.e. \( C = \{c_1, \ldots, c_n\} \), and a finite, non-empty set \( A \) of arcs which represent the relationships between those concepts (Albert & Steiner, 2005b). Every arc is an ordered pair from the set of concepts and is characterised by a relation label describing the relationships existing between those two concepts. Such a combination of two concepts and the labelled relation between them constitutes a proposition; i.e. a statement forming an elementary unit of declarative knowledge (e.g. Anderson, 1995; Ruiz-Primo, 2000). This means, a concept map is basically a representation of declarative knowledge by a collection of propositions. Most commonly, a concept map is depicted by a visual graph representation (see Figure 1 for an example concept map describing what a concept map is), although also other representation formats (e.g. proposition list, matrix) are possible. Concept maps are used in traditional and technology-enhanced learning in versatile ways (see Steiner et al., 2007) and approaches for the automated generation of concept maps for educational purposes are increasingly researched (Chen, Kinshuk, Wei, & Chen, 2008; Lau et al., 2009; Lee & Segev, 2012).

A crucial issue, once a domain ontology has been built either manually or automatically, is its evaluation. A specific ontology can be considered as an abstract and simplified view of the domain in question that is intended to be represented for a certain purpose (Gruber, 1995). Thus, a domain ontology and, respectively, a concept map constitute models of the current knowledge about the world for a given knowledge domain in a given context. For one and the same knowledge domain it is not realistic to assume that only one correct representation of complete consensus exists. Rather,
there is a range of alternative ways conceivable to describe and conceptualise the same domain (e.g. Brank, Grobelnik, & Mladenic, 2005). This is, because describing and structuring a domain necessarily entails some sort of world view or opinion (Kennedy, McNaught, & Fritze, 2004; Uschold & Gruninger, 1996). Especially in the humanities there may be different ways to look at issues, multiple viewpoints, or several answers to a problem. Moreover, the description of a domain may also differ due to its intended purpose and ultimate use. Assume, for instance, a domain ontology on basic arithmetics that is represented by a concept map. Such a concept map may be dedicated to different purposes. One application might be the use for teaching, i.e. for presenting learning content to students. Another conceivable application would be to convey preservice teachers the didactics of teaching basic arithmetics. For both purposes a concept map on the same domain will be needed, but they will necessarily picture somewhat different knowledge elements.

Since a body of knowledge may be conceptualised in many different ways and by different ontologies, approaches are needed that assist in deciding which of them suits a given purpose or predefined criterion best, or to decide whether a specific ontology is suitable for a specific purpose (Brank et al., 2005). Current research that elaborates on the evaluation of ontologies has proposed different evaluation approaches or frameworks and specifies a range of quality criteria, like for instance coverage, consistency, completeness, computational efficiency (e.g. Gómez-Pérez, 2001; Obrst, Ceusters, Mani, Ray, & Smith, 2007; Vrandecic, 2010; Zouaq & Nkambou, 2009). Depending on the type of ontology, the methodology that has been used for building it, its representation format/language, the discipline and intended application, and the situation when an evaluation becomes necessary, different evaluation methods will be relevant and applicable (e.g. Brank et al., 2005; Gómez-Pérez, 2001; Pammer, Scheir, & Lindstaedt, 2006). As a result, there is no single or generally established and accepted methodology for ontology evaluation. Rather, as ontologies are built and used in different contexts and for different purposes and depending on the aspect of an ontology that is considered for evaluation, diverse kinds and methods of evaluation will be needed. Existing evaluation methods, however, largely focus on the formalisation stage of ontologies and on performing logical inferences, which are usually not directly relevant for concept maps that are mainly intended as communication instruments (e.g. for learning content presentation) and in most cases will not be formalised and machine-readable. Almeida and Barbosa (2009) underline the importance of evaluating the content of an ontology and not solely technical issues. Need for further research in ontology evaluation is particularly seen in this direction, as well as in the development of user- and application centred evaluation methods (e.g. Gómez-Pérez, 2001; Tartir, Arpinar, & Sheth, 2010). Research on the evaluation of concept maps has been concentrated on evaluating individual concept maps as instruments for knowledge diagnosis (e.g. Ruiz-Primo & Shavelson, 1996) instead of evaluating concept maps representing domain ontologies. Systematic approaches for assessing a concept map with respect to topological aspects, i.e. its structure, have been developed (Valerio, Leake, & Cañas, 2008). Semantics, i.e. the content of a concept map, has been acknowledged as the more important component, which is at the same time less intensely

---

Figure 1. Concept map on what a concept map is.

Source: Adapted from Steiner et al. (2007).
researched—due to the higher complexity of evaluating it (Miller & Cañas, 2008; Valerio et al., 2008). The methodology demonstrated in this paper is centred on lightweight domain ontologies that are mostly manually controlled and represented in human language, exemplified for concept maps. The methodology is presented considering (but not limited to) an application in an educational, instructional context.

In Albert and Steiner (2005a, 2005b) and Albert, Steiner, and Heller (2006) methodological considerations on the evaluation of concept maps have been presented. The purpose of the present paper is to illustrate these theoretical considerations by an empirical study practically applying and testing the proposed validation methods. The remainder of this paper is structured as follows: Section 2 elaborates on the validity of concept maps, specifying the aspects of validity that are considered in our evaluation methodology and putting it in relation to related work. The concepts of content validity and application validity are described and the suggested approaches for investigating these validity types are summarised. After that, an empirical study is presented in Section 3 as an illustration of the proposed methodology for an example concept map on elementary geometry. The paper ends with a general discussion on the work presented (Section 4) and a conclusion highlighting issues for further research (Section 5).

2. Validity of concept maps representing domain ontologies

Given a particular purpose, a concept map representation is needed that integrates the knowledge of the domain in question and presents a common understanding, summarising possibly differing perspectives. Take for example an e-learning environment: To ensure that people are able to successfully learn from the system, it has to be based on a reliable and valid representation of the respective knowledge domain that is in line with the users’ personal understandings. This means, for an effective application of a concept map depicting a domain ontology a proof of its well-foundedness and applicability is a critical precondition. This is even more important when making use of automated approaches for extracting and creating concept maps, e.g. for later usage for instructional purposes (e.g. Lee & Segev, 2012; Zouaq & Nkambou, 2009). This calls for a theoretical framework for evaluating the quality of a concept map on a specific domain. To this end, subjective approaches of stating face validity and utilising ranking and tagging mechanisms for concept maps (e.g. in the context of concept map repositories, where people can share and publish their maps), even if provided by domain experts, are not sufficient and may only give a first indication on the quality of a concept map. Rather, there is an urgent need for systematic and sound procedures to more formally evaluate the quality of a concept map in terms of validity. Only after having proven the validity of a concept map it can be confidently applied. Note, that here we focus on the evaluation of concept maps in terms of domain ontologies, i.e. representing expert knowledge for diverse purposes of communicating or teaching information (e.g. learning material, navigation aid in e-learning, visualisation of complex ideas and interrelations ...). This research is generally different from approaches primarily intended to be used for scoring concept maps of individuals for knowledge assessment purposes, which is an issue that has been dealt with already more extensively in the literature (e.g. Novak & Gowin, 1984; Ruiz-Primo & Shavelson, 1996; Valerio et al., 2008). Such approaches may be adopted for a kind of criteria-based evaluation of a domain concept map with respect to different aspects. Oftentimes, though, such scoring rubrics focus too much on structural aspects and are in any case insufficient for giving evidence of the validity of a concept map representing a knowledge domain.

When considering the validity of concept maps, two types can be distinguished: content validity and application validity (Albert & Steiner, 2005a, 2005b). To ensure the wellfoundedness of a concept map, prior to its use both validity aspects should be considered. Thus, a concept map needs to be validated before it can go into action. In the sequel, a concept map that is to be validated is alternatively also denoted as target map (or target concept map).
2.1. Content validity

2.1.1. The notion of content validity
Content validity refers to the correct building of the content of a concept map, i.e. to the question whether it actually and accurately represents the knowledge of the domain of interest. It has to be determined whether the concept map constitutes a valid model of a part of the current knowledge about the world (e.g. Albert & Steiner, 2005a, 2005b; Almeida, 2009; Gómez-Pérez, 2001). The notion of content validation is also in line with the idea of “semantic evaluation”, i.e. the detection to what degree a created ontology reflects the knowledge of a domain, as outlined by Zouaq and Nkambou (2009). Content validity refers to the vocabulary and the taxonomic and semantic relations and can therefore be considered as an approach of evaluating the lexical/vocabulary level as well as the hierarchy and semantic relations of a domain ontology (Brank et al., 2005). Of course, the evaluation whether a concept map adequately reflects the knowledge elements of the respective domain will also need to take into account its intended purpose. Therefore, given the intended purpose or ultimate use of a concept map, content validity refers to the question whether it adequately reflects some kind of common understanding of the domain (Albert et al., 2006). This is in line with the above-described aspect of potentially alternative conceptions of a knowledge domain and the necessity of integrating them.

2.1.2. Validation approach
For examining the content validity of a concept map, empirically collected concept maps representing personal knowledge (in the sequel denoted as “criterion maps”) may be utilised as a criterion and basis for comparison (Albert & Steiner, 2005a, 2005b; Albert et al., 2006). These criterion maps are collected from individuals of different knowledge level, in order to gather a comprehensive picture of personal understandings of the domain. This means, instead of comparing the target concept map to one “gold standard”, as proposed in ontology evaluation research (e.g. Brank et al., 2005; Hazman et al., 2011), in our approach the conceptualisations that a range of (non-expert) individuals have of a given domain are used as a reference for comparison. The individual concept maps collected through a concept mapping task capture the persons’ ontologies of the domain in question (Cimolino, Kay, & Miller, 2004). While Almeida (2009) in his proposal for evaluating the content of an ontology refers to the assessment of experts, our validation approach aims at involving the whole knowledge range—from rather beginners to (nearly) experts—in the validation and to gather their personal ontologies. Several techniques are suitable for collecting these criterion maps, when the validity of a target map is to be examined (for an overview on different types of concept mapping tasks see Steiner et al., 2007). Among them are common methods, such as requiring individuals to create a concept map from scratch representing their understanding of the domain (i.e. map creation; see e.g. Ruiz-Primo, 2000), asking individuals to fill in partially blanked concept maps (i.e. map completion; see e.g. Schau, Mattern, Zeilik, Teague, & Weber, 2001), but also more recently proposed techniques such as presenting the propositions of a concept map as a correct-incorrect discrimination task (Steiner, 2004; Steiner et al., 2007). These different techniques considerably vary in the extent of constraints imposed and information provided to individuals. Which method is the most appropriate one for gathering criterion maps will therefore heavily depend on the particular validation objectives and requirements, but also on constraints regarding available time and budget. It may, for instance, be the case that not the whole target map, but only special parts of it are to be validated; accordingly, different methods for posing a concept mapping task will be suitable.

By comparing the target map with the collected criterion maps, i.e. by analysing the similarity between them, content validity can be examined. A high similarity between the target map and the criterion maps indicates a high degree of content validity of the target concept map. This essentially corresponds to the general idea of ontology evaluation through alignment with another ontology or gold standard (Obrst et al., 2007). Those approaches of investigating the similarity or equivalence, however, often only consider the conformance with respect to the concepts represented. When addressing the level of taxonomic and semantic relations, the consideration of the directionality of relations might be difficult (Brank et al., 2005). In our conception of content validity, the similarity of a target map to
collected criterion maps is primarily investigated on the level of propositions as elementary knowledge units, and thus incorporating the concept as well as the relational level, including the direction of relationships. It is, however, also possible to determine content validity by only addressing the concepts or the relationships represented in a target map.

On principle, for the comparison between target and criterion maps similarity measures as proposed in the context of comparing student-generated maps with an expert concept map for knowledge assessment purposes can be utilised (e.g. Chang, Sung, Chang, & Lin, 2005; Goldsmith, Johnson, & Acton, 1991; Ruiz-Primo, 2000; Takeya, Sasaki, Nagaoka, & Yonezawa, 2004). Furthermore, traditional similarity measures and measures of association for binary data, can be applied (e.g. Goodman & Kruskal, 1979). “Precision” and “recall” are mentioned by Brank et al. (2005) as possible measures for comparing an ontology with a gold standard; these measures can be adopted for our approach, as well. In essence, the similarity measures are based on $2 \times 2$ contingency tables—in the present context detailing for the target map and one criterion map the number of propositions that are contained in either both or in only one of them. Alternatively, the number of common and unique concepts or common and unique relationships may be used for deriving those measures—if only those special parts of a concept map are to be validated. The objective is to determine and evaluate shared and distinct elements of the two concept maps. In this context, however, symmetric measures seem to be not so well suited for appropriately describing similarity. This is, because it should be taken into account that a criterion map possibly refers to only a specific part or substructure of the target map, i.e. it may be the case that an individual concept map only covers partial knowledge of the domain. Elements that are contained in the target map but not in the criterion map should therefore not negatively affect so much the similarity measure, but rather those elements that are only contained in the criterion map (Albert et al., 2006). Such an approach can be realised by applying Tversky’s contrast model (Tversky, 1977).

On the whole, there are a variety of similarity measures that can be used for examining the similarity between criterion maps and a target map. These measures differ in the map characteristics they are focusing on and thus may lead to differing results. Until now, the theoretical considerations do not suggest a single similarity measure as best-suited for the purpose of content validation. Therefore, it appears desirable to apply a collection of selected similarity measure in an empirical example, for the purpose of comparison and in order to be able to learn more about the significance and explanatory power of different measures.

2.2. Application validity

2.2.1. The notion of application validity

The second type of validity, application validity, refers to the practical usability and usefulness of a target concept map. It is sometimes argued that the best way for evaluating an ontology is the application for which it has been created (e.g. Leclère, Trichtet, & Furst, 2002). Application validity refers to the question whether a concept map serves the purpose for which it has been designed. This conception therefore corresponds to the general idea of task- or application-based evaluation, i.e. investigating how useful an ontology is in a certain application scenario and how good the results that are produced in the application are (Brank et al., 2005; Hazman et al., 2011). For analysing this type of concept map validity, relevant situational performance is utilised (Albert & Steiner, 2005a, 2005b). Situational performance in this context means behaviour in real-world situations of a map’s application scenario that does not consist in performing a concept mapping task, such as, for example, problem solving, answering questions, or even social behaviour in given situations. Our approach addresses human behaviour and performance, as opposed to system behaviour and performance that is oftentimes targeted in other ontology evaluations (e.g. Obrst et al., 2007). As naturally a person’s understanding of a domain is reflected in his/her behaviour and performance in given situations, situational behaviour constitutes a suitable criterion for validation. Of course, for validating a target map in this way, a type of situational performance has to be chosen, that is closely linked to the purpose and intended application of the respective concept map. Depending on the purpose (e.g.
presenting learning material, describing social skills) that is foreseen for a target map, different kinds of situational performance (e.g. problem solving, behaviour in social situations) will be appropriate. If, for instance, a concept map has been created for an instructional purpose of communicating and teaching learning content on a certain knowledge domain, performance on typical questions or problems in the given domain seems an appropriate situational performance for investigating the application validity of this map.

Very concretely, for examining application validity of a concept map, Albert and Steiner (2005b) have suggested an approach utilising Knowledge Space Theory. This approach assesses the target map’s ability to predict relevant situational performance. After introducing the basic notions of the theoretical framework of Knowledge Space Theory in the sequel, the suggested validation approach is sketched.

2.2.2. Knowledge Space Theory
Knowledge Space Theory (e.g. Albert & Lukas, 1999; Doignon & Falmagne, 1985, 1999; Falmagne & Doignon, 2011; Falmagne, Koppen, Villano, Doignon, & Johannesen, 1990) provides a formal framework for structuring a domain of knowledge and for representing knowledge based on prerequisite relationships. A knowledge domain is characterised by a finite, non-empty set \( Q \) of problems. The knowledge state of a learner is represented by the subset of problems that he or she is capable of solving. Due to prerequisite relationships among the problems of a domain, not all subsets of problems are expected to be observable knowledge states. If two problems \( a \) and \( b \) are in a prerequisite relation, denoted by \((a, b) \in R\), from a correct solution of problem \( b \) the mastery of problem \( a \) can be surmised. In other words, problem \( a \) is a prerequisite problem for problem \( b \). Assume, for example, two problems of basic algebra, one requiring to add variables and the other one a linear equation. The first problem will be a prerequisite for the second one. In other words, being able to solve the equation will entail being also able to solve the addition. Such a prerequisite relation can be depicted by a Hasse diagram (see Figure 2 for an example), where descending sequences of line segments indicate a prerequisite relationship.

According to the prerequisite relation illustrated in Figure 2, from a correct solution of problem \( b \) the correct solution of problem \( a \) can be assumed, while the mastery of problem \( e \) implies correct answers to problems \( a, b \) and \( c \). The collection of knowledge states corresponding to a prerequisite relation, including the empty state \( \emptyset \) and the whole set \( Q \), constitutes the so-called knowledge structure \( K \). The knowledge structure corresponding to the prerequisite relation shown in Figure 2 is given by

\[
K = \{\emptyset, (a), (c), (a, c), (a, b), (a, b, c), (a, b, d), (a, b, c, e), (a, b, c, d), Q\}.
\]

The possible knowledge states are naturally ordered by set-inclusion, as can be seen in Figure 3. Given a knowledge structure, there are various possible learning paths from the naïve knowledge state (empty set \( \emptyset \)) to the knowledge state of full mastery (set \( Q \)). The knowledge structure depicted in Figure 3 suggests to present learning objects related to problem \( a \) (or, equivalently, \( c \)), first. Subsequently, material related to problems \( b \) or \( c \) (\( a \), respectively) should be presented, and so on. In Figure 3 one possible learning path is indicated by arrows describing the successive steps of the learning process. Thus, given the knowledge state of a learner, a knowledge structure provides useful information, which learning content should be presented next, but also which previously learned material...
should be reviewed (see e.g. Falmagne, Cosyn, Doignon, & Thiery, 2006). Furthermore, a knowledge structure builds the basis for efficient adaptive knowledge assessment that allows to determine the knowledge state of a learner by presenting only a subset of problems. Through exploiting the structure inherent to the knowledge domain and taking into account previous answers of an individual, only a subset of problems has to be presented (Dowling & Hockemeyer, 2001; Falmagne et al., 2006).

Competence-based extensions of Knowledge Space Theory incorporate competencies or skills into the theoretical framework (e.g. Düntsch & Gediga, 1995; Falmagne et al., 1990; Heller, Steiner, Hockemeyer, & Albert, 2006; Hockemeyer, Conlan, Wade, & Albert, 2003; Kickmeier-Rust & Albert, 2010; Korossy, 1997). These extensions aim at theoretically explaining the observed behaviour by considering underlying cognitive constructs. The main idea of these approaches is to assume a basic set of (elementary) competencies or skills describing abilities that are required for solving problems or that are taught by learning objects of a particular knowledge domain.

Knowledge Space Theory and its competence-based approaches have been successfully applied in the context of technology-enhanced learning for realising personalised learning paths and adaptive assessment (e.g. Albert, Hockemeyer, & Wesiak, 2002; Conlan, O’Keeffe, Hampson, & Heller, 2006; Kickmeier-Rust, Mattheiss, Steiner, & Albert, 2011). This theoretical framework enables the creation of tailored learning experiences that are characterised by an appropriate level of challenge for the learner and by didactically meaningful learning sequences based on the consideration of a knowledge domain’s inherent structure (e.g. Heller et al., 2006).

2.2.3. Validation approach
Assume a concept map representing a particular knowledge domain and for which application validity is to be examined. For testing the target map’s ability to predict relevant situational behaviour, problem solving has been identified as an appropriate measure of performance to be used as validation criterion. To this end, a set of typical problems is chosen, representing the domain in the sense of Knowledge Space Theory. For each problem the declarative knowledge that is required for solving the respective problem is determined by associating the problem with a substructure of the target concept map. This means, each problem is mapped to the concept map, by identifying the subset of propositions that represents the knowledge necessary for mastering the respective problem (Albert & Steiner, 2005b; Steiner & Albert, 2008). Each proposition might thereby be understood as kind of an atomic skill in the sense of competence-based extensions of Knowledge Space Theory. Based on the problems’ representation by substructures of the concept map, dependencies between problems in terms of a prerequisite relation can be derived by set inclusion (Albert & Steiner, 2005b; Steiner & Albert, 2008). Assume, for example, a problem $X$ that is represented by the proposition set \{P3, P7, P8, P9, P13, P15\} of a target concept map, and another problem $Y$ that has been associated with propositions \{P3, P7, P9, P13\}. As the representation of problem $Y$ constitutes a subset of that of problem $X$, it is assumed that $Y$ is a prerequisite for $X$. The dependencies derived in this way serve for
establishing a knowledge structure that collects the set of possible knowledge states. The knowledge states constitute answer patterns that are expected to be observable, provided that the target map validly represents the domain. The next step in the validation approach is therefore to collect empirical answer patterns on the set of problems. It can then be investigated whether the observed answer patterns correspond to the predicted knowledge states, by using a discrepancy index describing the similarity between the knowledge structure and the set of answer patterns (e.g. Doignon & Falmagne, 1999, chapter 12). As the target map has been used for establishing the knowledge structure, the empirically obtained answer patterns serve as validation criterion. If the empirical answer patterns correspond well to the predicted knowledge states, the concept map can be considered to be valid—assuming that both, the chosen set of problems as well as the sample of persons are adequate and representative.

Please note that situational behaviour and the according “problems” in this validation approach are not confined to performance measures and problems or questions known from traditional education, like math problems. Since Knowledge Space Theory and its competence-based extensions are applicable for items (and skills) in most diverse domains, like chess (e.g. Schrepp, Held, & Albert, 1999), moral thinking (Albert & Pillerstorfer, 2007), drumming (Rappitsch, Stamov-Roßnagel, & Albert, 2004), intercultural competence (Albert, Pivec, Spörk-Fasching, & Maurer, 2003), etc., the problems used for validating of a concept map may be of highly diverse nature, depending on the knowledge domain and the situational behaviour that is most suitable for application validation. The term problem therefore has to be understood in a wider sense in terms of problem situations.

3. An empirical study illustrating the validation methodology

The methodological considerations distinguishing two different types of validity for a concept map and the approaches for investigating them, as described above, have been applied in an empirical context in order to exemplify the validation methodology and to examine and demonstrate its significance and usefulness.

3.1. Overall method

3.1.1. Creation of a target concept map

For this empirical investigation the knowledge domain of elementary geometry, more precisely a small subdomain on right triangles and the theorems in right triangles has been chosen and considered from a teaching/learning perspective. Our intended aim has been to evaluate a concept map in terms of identifying whether it accurately presents the domain and fits the purpose of being used as teaching material in a high school maths course, i.e. presenting the learning content of the knowledge domain in question. A concept map has been created based on maths textbooks (e.g. Lewisch, 1988) as well as based on typical geometry problems in order to describe the knowledge of the domain and to serve as the target concept map for validation. The relevant learning contents from the textbooks have been extracted and broken down to propositions representing elementary knowledge elements of the domain. For representative geometry problems the steps for solving them have been analysed in order to identify declarative knowledge elements required and underlying the correct solution. The coverage of the respective propositions in the proposition set derived from the textbooks has been checked and in case of missing knowledge elements the respective propositions have been added. As the problems analysed include and require also knowledge on tangent lines of circles, the according knowledge elements have also been taken into account. The list of propositions derived in this way has served as a basis for the creation of a concept map for the selected knowledge domain. Thereby, one and the same concept appearing several times in the extracted propositions has naturally been represented by only one node in the concept map. Please note that the concept map has been created in German language and based on Austrian textbooks and curriculum and has been translated into English only in a subsequent step. The resulting target concept map contains 45 propositions consisting of two concepts (or, respectively, of three concepts—in case of relations with two end nodes) linked by one relationship. Figure 4 illustrates an extract of the concept map; the full list of propositions of the concept map can be found in Appendix A1.
3.1.2. Subjects and study setting

The derived target concept map has been examined in an empirical validation study. The aim of this investigation has been to analyse content validity of the concept map, on the one hand, as a critical precondition of the effective use of a concept map. On the other hand, application validity has been investigated in the same study, in order to prove the practical usefulness of the generated concept map for its designated purpose. Data collection for both aspects of validity has been carried out in single investigations. In total 44 subjects (20 male and 24 female) have taken part in the study. The participants have been 18 to 59 years old ($M = 26.23$, $SD = 8.77$). In order to balance eventual knowledge-activating effects, the order of data collection for content and application validity, respectively, has been balanced over the subjects. Randomly assigned to one group, from half of the participants first data concerning content validity has been gathered, and for the other half data collection referring to application validity has constituted the first part of the investigation. The duration for a complete single investigation has been about two hours ($M = 125$ min, $SD = 20.13$).

3.2. Examining content validity

3.2.1. Material and procedure

To examine the content validity of the target map individual concept maps (i.e. criterion maps) have been collected by posing a concept mapping task. This concept mapping task has been comprised of a correct-incorrect discrimination task (Steiner et al., 2007), which consists in presenting the propositions of a concept map together with distractors (i.e. incorrect statements) and letting the subjects judge the correctness of each statement.

The target concept map consists of 45 propositions, which make up the supposedly correct statements characterising the knowledge domain. In addition to these propositions 36 distractor items have been constructed also referring to the knowledge domain of right triangles but depicting assumingly invalid knowledge elements. Distractor statements have been built by combining concepts and relationships existing in the target concept map in different ways, such to form new propositions that represent invalid knowledge elements. Furthermore, distractors have been created by substituting concepts and/or relationships of valid propositions by new concepts or relation labels that are not part of the target concept map. In the construction of distractor items it has been tried to include potential or systematic misconceptions that a person might have with respect to the knowledge domain in question. In total 36 suitable distractor items have been derived (see Appendix A2 for the complete list).
All 81 statements (i.e. propositions from the target concept map and distractor items) together have been presented in randomised order to the participants. To make clear that the statements relate to right triangles, some of the statements have been complemented by including an explicit reference to the context of right triangles. Participants have been requested to judge each statement as either “correct” or “incorrect”. To ensure conscious and careful response behaviour and to avoid guessing, additionally a confidence rating has been collected for each statement, i.e. participants have been asked to indicate their confidence on the given judgement on a three-point rating scale (see Table 1 for an exemplary extract of the proposition correct-incorrect discrimination task). The items have been presented in a test booklet, whereby the single pages with their respective subsets of statements have been randomised for each participant. There has been no time limit for working on the proposition correct-incorrect discrimination task; the average duration for completion has been 27 min (SD = 6.68).

The collection of statements that have been judged by an individual as “correct” make up the individual concept map of this person, representing his/her personal understanding of the knowledge domain. Hereby, of course the variability of the individual concept maps is constrained to the statements presented in the task. In any case, this type of concept mapping task has ensured a smooth and manageable data collection, which has been necessary to enable data gathering for application validity in the same investigation. A data collection for both validity types in the same session has been required in order to ensure that participants have not been able to learn in between the sessions, thus realising equal conditions for both validation approaches.

The collected individual concept maps are used as criterion maps for content validation of the target map. In the context of the present investigation the aim is to validate the target map as a whole. Therefore, propositions of the target and criterion maps are considered for data analysis.

For giving evidence of the content validity of the target concept map similarity measures are calculated as proposed in Albert and Steiner (2005a, 2005b) and Albert et al. (2006). As there is a range of similarity measures available that seem appropriate for the purpose of such a validation; in the present empirical investigation several different measures are applied. This is, as one aim of this work is also to compare different measures in order to identify their expressiveness and to get an idea which measures are more or less strict or appropriate.

### 3.2.2. Results

In the correct-incorrect discrimination task on average 72.89 (SD = 2.62) out of in sum 81 statements have been judged correctly, with a range from at minimum 49 to at maximum 81 correctly judged items. Two people have a score of 81. The high resulting scores indicate that there is a quite high correspondence between the individual concept maps and the target map. This means, that to a large part propositions of the target concept map are also included in participants’ personal
understanding of the domain, while distractor items are largely not part of individuals’ conceptualisation of the domain. As a result, the ceiling effect that can be identified indicates actually a desirable and encouraging result and serves as an initial indicator for the validity of our target concept map.

For the purpose of the present paper in particular the appropriately judged correct statements, are of special interest for data analysis. In other words, it is analysed how many of the (in sum 45) target map’s propositions are also contained in the criterion maps. In terms of signal detection theory (Green & Swets, 1966) the respective set of propositions can be called “hits”. The criterion maps contain on average 40.7 (SD = 3.28) propositions of the target map (out of 44 possible). Furthermore, the number of propositions not covered in the criterion maps but being part of the target map (i.e. “misses” in signal detection theoretical terms) is of interest, which is on average 4.3 (SD = 3.28). Another relevant value for data analysis is the number of propositions that are part of the criterion maps but not contained in the target map (i.e. “false alarms”—these are on average 3.3 (SD = 3.11). Actually, these false alarms constitute distractor items that have been judged as being “correct”. Finally, the number of distractor items that have been correctly rejected (i.e. “correct rejections”) is on average 32.7 (SD = 3.11).

Each criterion map, i.e. each empirically collected individual concept map, is compared to the target concept map and the similarity between them is investigated. This is done by creating contingency tables detailing the number of propositions contained in both maps (hits), in only one of them (misses and false alarms), and in none of the two maps (correct rejections)—for an example see Table 2. As can be read off from Table 2, for this specific criterion map of one of the study participants there are 40 propositions contained in the target map that are also part of the criterion map, i.e. the respective criterion map features 40 hits. Five propositions that are contained in the target map are not covered by the criterion map (misses), and four false alarms—i.e. propositions that are part of the criterion map but not of the target map—have occurred. From the distractor items in the concept mapping task 32 have been appropriately rejected as “incorrect”.

Based on these tables the similarity measures are calculated for each criterion map (i.e. comparing it to the target map) and averaged over all criterion maps to yield an overall measure on content validity. The resulting similarity values range between 0 and 1, with a result close to 1 indicating a high similarity between criterion map and target map—and thus, arguing for the target map’s content validity.

For data analysis, first a measure that incorporates all cells of the contingency table in its calculation is used. The traditional similarity measure “Simple Matching Coefficient” is given by $SMC = \frac{(a + d)}{(a + b + c + d)}$ (Sokal & Michener, 1958, as cited in Bortz, 2005). For the SMC an average similarity between the target and the criterion maps of 0.906 (SD = 0.07) results, giving evidence of a high similarity. As another measure for investigating the similarity between the maps, the so-called closeness index proposed by Goldsmith et al. (1991), is chosen. Initially suggested to measure the similarity between students’ concept maps and a teacher or expert concept map, this index can be adopted to compare criterion maps to a target map for the purpose of content validation. The index refers to propositions, i.e. considers concepts and their links to each corresponding concept. Adapted

| Target map | Propositions contained | Propositions not contained |
|------------|------------------------|-----------------------------|
| Criterion map | Propositions contained | \(a = 40\) | \(b = 4\) |
|            | Propositions not contained | \(c = 5\) | \(d = 32\) |

(a–d) constitute the cell identifiers for indicating the formulas of the similarity measures.
to the notations used in the contingency table of Table 2 the closeness index (CI) is calculated by:

\[ \text{CI} = \frac{a}{a + b + c} \]

This measure conforms to the Jaccard (1908, as cited in Bortz, 2005) similarity measure (JS). By the calculation of this measure actually the focus is put on the conformance between target map and criterion map in terms of common elements. The average similarity index calculated for the 44 criterion maps is 0.848 (SD = 0.1), indicating a high proportion of similarity between the criterion maps and the target map.

In addition, the so-called “Convergence Score” and “Salience Score” as suggested and applied by Ruiz-Primo (2000) and Ruiz-Primo, Schultz, Li, and Shavelson (1998) for scoring student concept maps are adopted and calculated. The convergence and the salience score actually also correspond to the measures of precision and recall as described by Brank et al. (2005) as potential similarity measures in ontology evaluation. The convergence score (i.e. precision), for our purpose, indicates the proportion of propositions in the criterion map that are also contained in the target map, out of all possible propositions in the target map, i.e. \( \text{ConS} = \frac{a}{a + c} \). The mean convergence score is 0.905 (SD = 0.07). The salience score (i.e. recall) indicates the proportion of propositions in the criterion map that are also contained in the target map out of all propositions in the criterion map, i.e. \( \text{SalS} = \frac{a}{a + b} \). A mean salience score of 0.925 (SD = .07) is found for the criterion maps. Both measures underline that most of the propositions contained in the criterion maps are also present in the target map.

Except for the simple matching coefficient all similarity measures do not take into account the “correct rejections”, i.e. cell “d” in Table 2. Actually, a number of propositions that is neither contained in the target nor in the criterion map can be indicated in our case of a correct-incorrect discrimination task. When using a different kind of concept mapping task, though, such as map creation or map completion, it will not be possible to indicate this number. The respective cell of the contingency table will be unallocated. In such case, therefore the SMC will coincide with the closeness index (and Jaccard similarity measure, respectively).

In a further step, Tversky’s contrast model (1977) is investigated into its use for content validation (see Figure 5 for a schematic illustration). In this way, an analysis may be realised that takes into account the fact that a criterion map may only cover a part (i.e. partial knowledge) of the target map. Following this consideration, propositions included in the target map \( t \) but not covered by a certain criterion map \( c \) (\( T-C \) in Figure 5) should not negatively influence the similarity measure. Conversely, propositions that are only contained in the criterion map but not in the target map (\( C-T \) in Figure 5) should reduce similarity. Hence, the similarity between a criterion map and the target map should not be viewed to be a symmetric relation. Rather, the directionality of comparison should be taken into account (Tversky & Gati, 1978). According to the contrast model the similarity between a criterion map and a target map is given by \( s(t, c) = \theta f(T \cap C) - \alpha f(T - C) - \beta f(C - T) \), where the terms \( \theta, \alpha \) and \( \beta \) reflect the weights given to the common and distinctive components. For our purpose the aim is not to investigate the degree to which target and criterion maps are similar to each other (nondirectional), but rather to investigate the degree to which the criterion map is similar to the target map (directional). Hence, it is assumed that the focus of attention is on the criterion map, such that its propositions will be more heavily weighted. In other words, those propositions unique to the criterion map are the ones that should asymmetrically affect the similarity computation; thus the value chosen for the parameter \( \beta \) is larger than \( \alpha \).

Figure 5. Schematic illustration of the relation between the target and criterion maps’ element sets.

Source: Adapted from Tversky and Gati (1978).
With respect to the weighting parameters, the following holds (Tversky, 1977): If \( \theta = 1 \) and \( \alpha = \beta = 0 \), the similarity between the two entities of comparison, i.e. between target and criterion map, is determined only by their common features. If, instead, \( \theta = 0 \) and \( \alpha = \beta = 1 \), then the similarity is only determined by their distinctive features. In the present context, as indicated above, a similarity estimation is desired that is influenced by the shared features of the criterion and the target map as well as by the propositions solely represented in the criterion map, whereas the propositions that are only part of the target map should not negatively affect similarity. Correspondingly, the following parameter values may be applied: \( \theta = 1, \alpha = 0, \beta = 1 \).

A generalised form of the contrast model is the so-called ratio model (Tversky, 1977). In this model, the similarity is given by

\[
S(t, c) = \frac{f(T \cap C)}{f(T \cap C) + \alpha f(T - C) + \beta f(C - T)}
\]

The ratio model defines a normalised value of similarity, such that the similarity value ranges between 0 and 1. As this normalised measure is more appropriate for comparison purposes, the ratio model has been applied for calculating the similarity in the present context. With the weighting parameters \( \alpha = 0 \) and \( \beta = 1 \), the calculation of similarity actually corresponds to the salience score presented above, i.e. a mean similarity of 0.925 (SD = 0.07) results according to the ratio model over all criterion maps.

It appears, however, reasonable to not completely disregard the propositions solely contained in the target map; rather, they may be taken into consideration, with a lower weight than elements that are unique to the criterion maps. This is of relevance if the validity testing shall take into account the aspect of "conciseness" in terms of whether a concept map is free of unnecessary, useless, or redundant elements (e.g. Gómez-Pérez, 2001; Vrandecic, 2010). When choosing slightly different weighting parameters, i.e. when also allowing the unique propositions of the target map to influence the similarity measure by increasing the value for \( \alpha \) from 0 to 0.3 (0.5, respectively), a slightly lower but still high average similarity value of 0.9 (SD = 0.08) results (\( M = 0.88, SD = 0.08 \), respectively, for \( \alpha = 0.5 \)), which still gives evidence of a high similarity between the criterion maps and the target map.

On the whole, the similarity measures calculated for the purpose of validating the target concept map on elementary geometry consistently show that the criterion maps are highly similar to the target map. This can be interpreted as giving evidence for the target map’s content validity. The concept maps of individuals (i.e. criterion maps) correspond well to the knowledge represented in the target map. The target concept map can be considered as a valid representation and knowledge model of the domain of interest, which is an important precondition for its use in instruction.

3.2.3. Discussion

This first part of our empirical study demonstrates the validation approach for investigating content validity on an example target concept map for elementary geometry. In the light of the analysis comparing the empirically collected criterion maps with the target map, the target map can be considered to be content valid. The personal ontologies on the knowledge domain gathered from persons of different knowledge level correspond well to the target map. Thus, the target map can be assumed to adequately reflect and appropriately represent a common understanding of the respective domain.

This encouraging result, however, may have been supported through the method applied for collecting the criterion maps. As the concept mapping task requires to judge predetermined statements as either correct or incorrect, the collected criterion maps are restricted to the predefined propositions (from the target map) and distractor items, and necessarily show a certain degree of similarity to the target map. This is even more relevant given the fact that the number of distractor
items presented is slightly lower than the number of propositions from the target map. In an optimal case, the same number of distractor items as propositions should be presented for such a correct-incorrect discrimination task—in the present context, however, attention has been paid to compose and provide distractors that are supposed to be actually probably part of individuals' personal understanding, because of some misconceptions or of some different kind of understanding, for instance. Thus, very unrealistic and far-fetched distractor items have been avoided. Generally speaking, a higher proportion of distractor items would most probably also lead to an increase of variability in the criterion maps, thus resulting in at least slightly lower similarity scores and consequently in a slightly less powerful evidence of the target map's content validity. Considering, however, a total number of correctly judged statements of 40.5 (in contrast to 81 on principle possible and on average 72.8 empirically observed) in case of random answers (with guess probability of 0.5 for each item), with a number of 22.5 hits (in contrast to on average 40.7 observed), it is obvious that the similarity between the collected criterion maps and the target map is significantly higher than in case of random answers (compare Table 3). Correspondingly, the similarity measures resulting for random answers are also considerably lower than measures derived for the empirical criterion maps (SMC = 0.38, CI = 0.5, ConS = 0.5 and SalS = 0.56 for random answers compared to values located around 0.9 as resulting for the empirical data in our study).

In addition, in our data collection on content validity, a careful answer behaviour has been ensured by requesting confidence ratings for each correct/incorrect-judgement. These judgements have not been considered in the data analyses presented above. Another possibility of analysis, though, would be to take into account only statements that have been judged with medium and high confidence, and to exclude those judgements of low confidence from data analysis. It has to be underlined, however, that this procedure somehow penalises subjects who—due to personality traits—are tempted to give rather unassertive judgements. When actually only taking into account judgements with a high confidence (“quite sure” and “very sure”) and disregarding all other elements, an even higher similarity between the criterion maps and the target map can be determined. For the simple matching coefficient a value of 0.93 results (SD = 0.05; compared to SMC = 0.9 when neglecting confidence judgements). The closeness index yields a similarity of 0.89 (SD = 0.77; compared to CI = 0.85) when taking into account confidence judgements, and the similarity calculated based on Tversky’s ratio model is 0.93 (SD = 0.06 with a = 0.3 and s(t, c) = 0.92, SD = 0.06 with a = 0.5).

Another issue with the applied concept mapping task is that the scores in the proposition correct-incorrect discrimination task are consistently quite high, i.e. a ceiling effect has been determined. First of all, this is not to be considered problematic, as this generally argues for the validity of the target map and can be used as an initial indicator for content validity. Having in mind the hypothesis that the target map is valid, actually high scores in the concept mapping task are expected. Given the participants of our study, who have been mainly recruited from the population of university students (of psychology, architecture and civil engineering) it has been expected to have a certain expertise in the knowledge domain. Although a sample with a certain level of knowledge in the respective domain is definitely a necessary precondition for carrying out a validation study, it may be worth considering the (additional) involvement of more inexperienced participants in future investigations (in our concrete case school children of different ages), such to more comprehensively cover different levels of expertise of the knowledge domain.

| Target map | Propositions contained | Propositions not contained |
|------------|------------------------|-----------------------------|
| Random answer criterion map | a = 22.5 | b = 18 |
| | c = 22.5 | d = 18 |
In our example, the validity analysis has been carried out for the whole target concept map, i.e. on the level of propositions, and for the full set of propositions. A benefit of the applied technique for collecting criterion maps is therefore that it also allows gathering data for validating only part of a target map. For instance, it might be necessary to validate only a specific substructure of the concept map. This could for example be the case, if the scope of an existing, already validated concept map is broadened through extension with respect to a certain subtopic. Then, only the respective amendment may be addressed in a validation study, thus presenting only the propositions of this part in the correct-incorrect discrimination task. Another aspect of validating only part of a concept map may be to investigate the validity of the relations of a target concept map that, presumably, contains an already validated set of key concepts of a domain. In this case, distractor statements in the correct-incorrect discrimination task would be characterised by simply using alternating or alternative relation labels. A validation of only the concepts of a target map might be of interest, e.g. if the validation aims in finding a valid set of concepts representing a domain, which can subsequently be provided to students in an educational context requiring them to create concept maps during or after learning. For the assessment of learning performance those maps could afterwards be compared to an expert concept map created by the teacher using the same set of concepts. Another reason to consider only concepts (or, alternatively, relations) for content validation would be in case of a multi-step evaluation approach already incorporated in the process of building a target concept map, i.e. when building and investigating the validity of a domain ontology in several iterative steps. In this sense, validation can be seen as a support for concept map creation and ontology engineering (e.g. Pammer et al., 2006).

3.3. Application validity

3.3.1. Material and procedure

For examining application validity the target concept map has to be related to situational behaviour in its intended use context. As mentioned before, the target map has been created having an application in an educational and learning context in mind. Assuming that the target concept map would later on be applied as instructional material in classroom to teach a subdomain on right triangles, problem-solving behaviour in the respective knowledge domain has been selected as an appropriate situational performance for validation purposes. To collect behavioural performance data, a set of in total 13 geometry problems has been adapted from Korossy (1993, 1996). These problems constitute typical and representative problems of the knowledge domain in question (see Figure 6 for an example and Appendix A3 for the full set of problems); three problems have been used only as warming-up exercises in the beginning of the investigation, the other ten problems have been presented in randomised order and have been utilised for actual data collection and analysis.

To investigate application validity, the target concept map is used for establishing a knowledge structure on the geometry problems (compare Section 2.2.3 and Albert & Steiner, 2005a, 2005b), i.e. predictions on problem-solving performance are derived from the concept map. This means, each of the ten geometry problems is mapped onto the target concept map by identifying the propositions representing the knowledge necessary for solving it. A subset of in total eleven propositions is identified to constitute fundamental knowledge of the domain that is required for all problems and their solutions. Correspondingly, the respective proposition set is associated with each problem. It has to be taken into account that most of the problems may be solved in several different ways. Thus, first the different solution ways for each problem are identified and collected based on cognitive task analysis and on the work of Korossy (1993). For every problem, each solution way is associated with the relevant propositions of the concept map. This means, a problem is represented not only one time on the concept map, but rather two (or more) times, corresponding to the different options of solving it. For the geometry problem presented in Figure 6, for example, two solution ways are applicable. Correspondingly, the problem is associated with two different subsets of propositions on the concept map, which are overlapping to a large degree but differing in parts according to the different solution strategies (see Table 4). The complete list of problems with concept map substructures (i.e. propositions subsets) assigned to their individual solution ways are documented in Appendix A4.
For deriving dependencies between the problems in terms of a prerequisite relation their representations are compared to each other by means of the set inclusion principle. More precisely, the proposition sets associated with problems’ solution ways are compared. In a first step, subset relations between solution ways are identified and documented. These are interpreted as preliminary prerequisite relationships among the respective problems. As there is more than one solution way for each problem, for few pairs of problems the comparison of certain solution ways indicates a prerequisite relationship in one direction (e.g. problem X is a prerequisite for problem Y), whereas the comparison of other solution ways indicates a prerequisite relationship in the inverse direction (i.e. problem Y is a prerequisite for problem X). Therefore, in a second step these preliminary assumptions are cleaned up by eliminating contradictory prerequisite relationships between problems, while retaining only definite ones. In a final step two additional prerequisite relationships (from the originally identified ones) have been added to maintain transitivity in the prerequisite relation (for the complete documentation of the steps for determining prerequisites between problems please refer to Appendix A5).

The prerequisite relation resulting for the ten geometry problems on the basis of their association with the target map is depicted in Figure 7. The knowledge structure corresponding to this prerequisite relation consists of 44 knowledge states. In other words, the established prerequisite relation restricts the number of possible answer patterns from 1,024 (i.e. 2^{10}) to only 44. For instance, each knowledge state containing problem d is assumed to also contain problem a, and a knowledge state covering problem g necessarily should also include problems a, c and f.

The knowledge structure induced by the established prerequisite relation and collecting the knowledge states expected to be observable, is presented in Figure 8. The knowledge structure derived in this way serves as a basis for investigating the target map’s application validity. The
knowledge states collected in the knowledge structure constitute answer patterns on the 10 geometry problems in the sense of theoretical predictions on problem-solving behaviour—assuming that the knowledge structure and the concept map on which basis it has been constructed are valid. In order to examine whether these theoretically expectable answer patterns predict empirical answer patterns well, the geometry problems have been presented to participants in our empirical study.

First, the three warm-up exercises have been presented, which are not taken into account for data analysis, but rather have served for making the subjects familiar with the type of problem-solving task and for reducing test anxiety. Subsequently, the ten geometry problems that have been mapped to the target concept map have been presented in randomised order. Each problem has been depicted on a separate sheet. Participants have been requested to work at least three minutes on a problem, at maximum ten minutes. This has ensured that a person being able to solve a problem also brought this ability out, on the one hand, and to limit the duration for the single investigations on the other hand. The average duration for working on the problems has been 66 min (SD = 12.38).

The answer patterns collected from the 44 participants are analysed for whether they correspond well to the theoretical knowledge structure derived from the concept map. The extent of conformance is interpreted as a measure for the target map’s application validity.

The adaptation of the geometry problems from Korossy has been very useful, as for the same ten problems a different knowledge structure has been established by Korossy (1993, 1996). In that prior work, though, instead of building a prerequisite relation on the problems a knowledge structure has been generated on the basis of an underlying competence modelling and assignment. The respective structure features only 25 knowledge states (see Figure 9), i.e. this structure is more restrictive. When comparing Korossy’s and our structure (see Figures 8 and 9), it is evident that problems c and a in both structures consistently resulted to be basic problems without any prerequisites. In both structures

![Figure 8. Knowledge structure established on the geometry problems.](image)
problem e stands out as a quite difficult problem with many prerequisites and being part of only higher order knowledge states in the structure. Also most other problems can be found at similar levels in the structures when considering the cardinality of knowledge states in which they appear. Problems b and h, though, are exceptions. While in the structure of Korossy problem b seems to be a rather basic problem that is solvable in knowledge states that are located quite at the bottom of the structure, in the structure resulting from the concept map this problem turns out to be rather challenging and appears first only in knowledge states with a cardinality of at least five problems. In the structure derived from the concept map, problem h occurs first in knowledge states with a cardinality of 3, while in Korossy’s structure this problem is only contained in knowledge states at a medium level, starting from a cardinality of 6. In total, there are 16 knowledge states (including the empty set and the set of all ten problems) that the two structures have in common (i.e. their intersection; see Figure 9). The structure of Korossy is compared to the knowledge structure derived from the target map; more precisely, the goodness of fit of both structures in terms of predicting the empirical data-set is investigated in data analyses.

3.3.2. Results
For evaluating the solutions of the geometry problems two different approaches are realised. On the one hand, a geometry problem is categorised as correctly solved if the solution way and the final result is correct (evaluation approach 1). In a second evaluation also those problems are considered as mastered, where the solution way is on principle correct, but a calculation failure has occurred,
thus leading to an incorrect numerical result (evaluation approach 2). From the ten geometry problems on average five have been solved correctly (with $M = 4.82$, $SD = 2.47$ for evaluation 2 and $M = 5.02$, $SD = 2.67$ for evaluation 2). The range of mastered problems is 0–8 for evaluation 1 and, respectively, 0–10 for evaluation 2. Two participants (independent of the evaluation approach) have not solved any of the geometry problems, and only one person has been able to master the full problem set in case of evaluation approach 2.

The solution frequencies for each individual problem provide a first indication on the fit of the empirical answers to the theoretically assumed prerequisite relation (see Figure 7). If for a pair of items $x$ and $y$ a prerequisite relationship is assumed according to the prerequisite relation in terms of $x$ is a prerequisite for $y$; then $y$ should only occur in an answer pattern if also problems $x$ has been solved. As a result, it is expected that for the overall sample the solution frequency for item $x$ should be equal or higher than for $y$. If empirical solution frequencies significantly contradict the prerequisite relation, this would indicate problematic and invalid assumptions. As can be seen in Figure 7, the empirical solution frequencies correspond very well to the theoretical prerequisite relation. Only for problem $c$ for both evaluation approaches a slightly lower frequency than for its successive, superordinated problem $g$ has been observed. The difference between the solution frequencies of problems $c$ and $g$ is, however, marginal—and this result is therefore not indicative for an invalidly assumed prerequisite relationship. In case of evaluation 1 problem $f$ results to be solved slightly too often, but again the contradiction to the theoretical structure is not very strong (i.e. difference is very small); and when considering evaluation approach 2 it even disappears.

For comparing the empirical answer patterns to the theoretical knowledge structure the minimal symmetric distances between empirical and theoretically expected answer patterns are calculated (Garnier & Taylor, 1992; Kambouri, Koppen, Villano, & Falmagne, 1994). The minimal distance ($d_{\text{min}}$) between an answer pattern and a knowledge structure is defined as the distance to the nearest knowledge state. Assume for example the answer pattern \{$a, d$\} for the five problems presented in the example in Section 2.2.2 (Figures 2 and 3). This answer pattern is actually not part of the knowledge structure, the nearest knowledge states are given by \{$a$\} and \{$a, b, d$\}, which both differ in one problem (i.e. $b$ or, respectively, $d$) from the answer pattern. Thus, the distance between the answer pattern \{$a, d$\} and the respective knowledge states is one (in case of state \{$a$\} in solving problem $d$ a lucky guess has been made or, respectively, in case of the state \{$a, b, d$\} on problem $b$ a careless error has occurred—assuming that the knowledge structure is correct). The minimum possible distance is given by 0, meaning that no deviation from the knowledge structure occurs, i.e. the answer patterns fully correspond to the predicted knowledge states of the theoretical knowledge structure. The maximum possible distance is given by half the number $n$ of items (i.e. $n/2$ for even, or, respectively,

| Table 5. Distance distribution and mean symmetric distances between the collected empirical answer patterns ($N = 44$) and the knowledge structure derived in the present investigation (based on the concept map) and, respectively, the knowledge structure as established by Korossy (1993) |
|-----------------------------------------------|
| **Knowledge structure derived from** | **Knowledge structure of Korossy (1993)** |
| **Frequency** | **Frequency** |
| **Distance** | **Evaluation 1** | **Evaluation 2** | **Evaluation 1** | **Evaluation 2** |
| 0 | 18 | 24 | 18 | 24 |
| 1 | 20 | 16 | 19 | 15 |
| 2 | 5 | 3 | 7 | 5 |
| 3 | 1 | 1 | 0 | 0 |
| Mean distance | 0.75 | 0.56 | 0.75 | 0.57 |
| (SD = 0.74) | (SD = 0.72) | (SD = 0.71) | (SD = 0.69) |
(n – 1)/2 for odd numbers of items). Hence, for the present investigation with 10 geometry problems the greatest possible distance is 5. The averaged minimal distance across all response patterns constitutes a measure for the correspondence between empirical data and theoretical knowledge states; it gives evidence of the validity of the established knowledge structure and thus, of the target map. In the present context, \( d_{\text{min}} \) is consequently a measure for the target map’s ability to predict relevant situational performance.

The average symmetric distance between the 44 empirically collected answer patterns and the hypothesised knowledge structure is 0.75 (SD = 0.74) for evaluation 1. For the second evaluation approach, i.e. judging problems with the correct solution way as being mastered (irrespective of probable calculation errors), the mean distance is with 0.57 (SD = 0.72) even lower (see Table 5). Our structure is also compared to the knowledge structure established by Korossy (1993; compare Figure 9). When calculating the fit of our empirical data to Korossy’s structure, surprisingly nearly identical results can be obtained: A mean distance of 0.75 (SD = 0.71) results for evaluation 1, and for evaluation 2 an average distance of 0.57 (SD = 0.69) can be determined. Table 5 presents in detail the distance distributions and mean distances for the two knowledge structures and evaluation approaches.

When calculating minimal distances between an empirical data-set and a theoretical knowledge structure, it has to be taken into account that trivial answer patterns (i.e. none or all problems mastered) do not provide any information of the validity of the specific theoretical structure, as these knowledge states are contained in any knowledge structure. In the present investigation, however, the number of trivial answer patterns is very low; in evaluation 1 there are only two answer patterns with no problem solved correctly (i.e. \( N = 42 \) non-trivial answer patterns), in evaluation 2 there is one additional answer pattern with all problems mastered (i.e. \( N = 41 \) non-trivial answer patterns). When considering only non-trivial answer patterns for the calculation of mean distances slightly higher distance values are obtained (for details see Table 6).

When having a closer look at the distances and nearest knowledge states for both knowledge structures, it can be determined that in each case 17 out of 18 answer patterns that have a distance of 0 actually correspond to knowledge states that are part of the intersection of both structures. For the majority of answer patterns featuring a distance of 1 (17 answer patterns) or 2 (4 answer patterns), the nearest knowledge state(s) also constitute states that are contained in both knowledge structures. This means, the empirical answer patterns collected actually do not allow to discriminate well between the structures and therefore lead to highly similar results. Although the structure of Korossy is the more restrictive one, it has to be underlined that the main aim in the present investigation has not been to derive a very compact and efficient structure, but to predict problem-solving behaviour on the basis of a concept map in order to investigate the map’s application validity. The results for the knowledge structure established in the present investigation for application validation are therefore very satisfactory.

The comparison between the empirical answers and the knowledge states predicted from the concept map yields definitely encouraging results and argues for the properness (more precisely, the

| Table 6. Mean symmetric distances resulting for the two knowledge structures when considering only non-trivial answer patterns |
|---------------------------------------------------------------|
| **Knowledge structure derived from concept map**              |
| **Knowledge structure of Korossy (1993)**                     |
| Evaluation 1 \( (N = 42) \)          | Evaluation 2 \( (N = 41) \)          | Evaluation 1 \( (N = 42) \)          | Evaluation 2 \( (N = 41) \)          |
| Mean distance | 0.79   | 0.61   | 0.79   | 0.61   |
| \( (SD = 0.74) \) | \( (SD = 0.73) \) | \( (SD = 0.71) \) | \( (SD = 0.69) \) |
validity) of the established knowledge structure. For a statistical test of the knowledge structure’s validity, Heller (2001) suggests to use the frequency distribution of the symmetric distances. The validity of the knowledge structure is estimated by using a one-dimensional $\chi^2$ statistic. The distribution of distances for the items’ power set (i.e. all answer patterns) is utilised as a basis for the test. It is examined whether the distribution of distances for the empirical answer patterns differs significantly from the distribution for the expected patterns. For the present investigation the null hypothesis (that the empirical data comprises no structure) can be rejected (at a 1% level of significance; $\chi^2_{0.01} = 15.08$ and $\chi^2_{obs} = 172.68$ for evaluation 1 and $\chi^2_{obs} = 289.26$ for evaluation approach 2). This result argues for the validity of the knowledge structure.

As another measure for investigating the validity of the theoretical knowledge structure, the distance agreement coefficient $DA$ (Schrepp, 1993; Schrepp et al., 1999) is calculated. The DA is a measure for the fit between a knowledge structure and an empirical data-set taking into account the size of the knowledge structure (i.e. the number of knowledge states). For this, the empirical average symmetric distance is relativised to the mean distance of the power set. A lower DA indicates a better fit of the knowledge structure to the data. For our investigation a DA of 0.35 (and, respectively, 0.26 for evaluation 2) resulted. As this measure takes into account the size of a knowledge structure, it is suitable for comparing the results to the structure of Korossy (1993), which yielded a DA of 0.32 (and 0.24 for evaluation 2, respectively). As can be seen, these measures are very similar for both structures, with marginally better results for the structure established in prior work.

To give a further indication of the quality of the results for the present study, a simulation is carried out for comparison with the empirically observed answer patterns. To this end a data-set is simulated, taking into account the solution frequencies for items, i.e. the number of persons having mastered a certain problem is equal in both, the empirical and the simulated sample. This can be called a “frequency simulation”, as specific information on solution frequencies is taken into account (Wesiak, 2003; chapter 3). Other possibilities for testing the empirical validity of a knowledge structure would be random simulations, probability simulations, or a resampling approach (Heller & Albert, 2005; Wesiak, 2003, chapter 3). The simulated sample shows an average minimum distance of 1.47 (SD = 0.90) and is utilised for calculating a statistical test in order to compare the distance distributions in the empirical and the simulated data-sets. The result is highly significant ($U = 357.5$, $p = 0.001$), indicating that the symmetric distances of the empirical answer patterns are significantly lower than those of the simulated answer patterns.

In sum, the data analysis leads to the conclusion that the empirically collected answer patterns on the geometry problems are predicted well by the theoretically expected knowledge states as derived from the mapping of the geometry problems to the concept map. Since the correspondence between the theoretical knowledge structure and empirical performance can be interpreted as a measure of application validity, the target map can be regarded as validated and is ready for use.

### 3.3.3. Discussion

The second part of our empirical study demonstrates the approach for application validation, exemplified for the example target concept map on elementary geometry. Problem-solving performance has been selected as a suitable behavioural and situational performance measure relevant for the intended application of the concept map, and a set of representative problems of the domain of interest has been adopted from a prior study. The target concept map has been used to predict answer patterns in problem-solving performance by assigning problems with subsets of the concept map and deriving prerequisite relationships for establishing a knowledge structure. The comparison between empirical answer patterns and theoretically expected knowledge states yields a good fit and thus, argues for the validity of the knowledge structure and its underlying concept map.

Contrasting the results obtained for our knowledge structure with the distances between the data and a different structure established for the same problem set in a prior investigation of Korossy (1993), very similar results can be determined. As the collected answer patterns, however, mostly
feature a minimal distance to knowledge states that are part of the intersection of both structures, the empirical data does not allow a good comparison and discrimination between the two knowledge structures. To make this possible the involvement of individuals showing a broader variance in expertise and, consequently, performance would be needed in future investigations, such to have a broader range in responses and thus, a more extensive coverage of possible answer patterns. From the perspective of Knowledge Space Theory, the structure as reported in Korossy (1993) can be argued to have a slight advantage over the structure established based on the concept map, as it is the smaller structure and the DA is marginally lower than for our structure. It has to be noted, however, that in case of the present investigation the aim has not been on generating a knowledge structure that is as small as possible, but rather to predict situational behaviour on the basis of a concept map in order to investigate its application validity. Therefore, the fit obtained in our investigation between empirical data and our theoretical model is considered highly encouraging and satisfactory.

Given that the target concept map focuses on theorems in right triangles, the representation of the geometry problems on the concept map only considers solution ways for the problems using the theorems in the right triangle (i.e. Pythagorean, Altitude and Euclidean Theorem). Actually, the geometry problems would also be solvable in somewhat different ways, by applying trigonometric functions or quadratic equations. As the target concept map represents only a subdomain on right triangles referring to the theorems in right triangles, those alternative solution ways are not presentable and not intended to be captured. Strictly speaking, for the purpose of application validation therefore only answer patterns should be utilised that actually use those solution approaches that have been addressed and used for establishing the knowledge structure. Hence, solution behaviour corresponding to alternative solution ways should be excluded from data analysis. In the present investigation, contrary to the instruction given in the beginning, eight out of the 44 participants partly have used alternative trigonometric solution approaches. When excluding the answer patterns of the relevant eight persons from data analysis and calculating the average minimal symmetric distance between the empirical data and the theoretical knowledge structure, the result changes indistinguishably. The average distance in this case is with $M = 0.75$ (SD = 0.76), actually the same result for the whole sample ($M = 0.75$, SD = 0.74).

The utilised performance measure actually involves also procedural knowledge to a quite high degree, as the solution of the geometry problems does not only require knowing the theorems in the right triangle on a solely factual, conceptual level, but rather also requires abilities in terms of procedural knowledge, i.e. on how to apply the declarative knowledge elements. The target map in this investigation has only covered declarative knowledge, such that the establishment of the knowledge structure has been based only on the problem’s declarative representation on the target map. The concept map’s predictive power has therefore been investigated using a more comprehensive situational behaviour and thus, an even stricter and more delicate validation criterion. Consequently, the results of the investigation are even more interesting and confirmatory for application validity. Hence, in line with the results of Steiner, Albert, and Sternad (2012) it has been demonstrated that it is possible to obtain a valid knowledge structure on procedural problems by representing them on a concept through solely the declarative knowledge relevant for them.

When nonetheless aiming at a closer alignment between target map and the represented problems a slightly different type of problems might be sought that refer to situational behaviour relevant for the target map’s intended purpose and that require largely declarative knowledge. In our example, for instance, questions querying factual knowledge elements on the right triangle (e.g. the Pythagorean Theorem, the sides of a right triangle ...) could be used. The representation of such problems on the target map is more straightforward, and it is possible to cover the whole extent of underlying and required knowledge. When thinking about different knowledge domains and target maps, though, in many cases situational behaviour will most probably involve also procedural knowledge to some extent. If a target concept map also covers procedural knowledge elements (e.g.
Pernaa & Aksela, 2010), a more exhaustive representation of problems, accounting also for proce-
dural aspects could be realised.

The target concept map has been created with teaching purposes in mind, i.e. to be used as learn-
ing material in a mathematics course. Naturally, the aim of teaching the respective subdomain of
gometry will be that learners not only acquire declarative knowledge, but also rather develop skills
to solve geometry problems requiring also procedural knowledge. According to the educational ap-
proach for explaining the linkage between declarative and procedural knowledge (e.g. Haapasalo &
Kadijevich, 2000), the development of procedural knowledge depends and builds upon declarative
knowledge. Correspondingly, in educational practice first a body of declarative knowledge needs to
be established before procedural knowledge elements can be taught and learned. In the context of
using concept maps for teaching, one possibility for supporting this development would be to try to
make explicit also procedural knowledge elements and include them in the target concept map (e.g.
Sims-Knight et al., 2004). In this way teaching could be assisted and the process of establishing and
linking knowledge could be supported through visually depicting and connecting declarative as well
as procedural information. The inclusion of procedural aspects in our target map would subsequent-
ly also allow a more comprehensive representation of the geometry problems for the purpose of
investigating application validity.

The problems, for which answer patterns have been predicted on the basis of the concept map,
are generally solvable through different solution ways. This complicates the derivation of prerequi-
site relationships among the problems for establishing a knowledge structure. As each problem is
represented by more than one proposition subset on the concept map, the prerequisite relation on
the problem set had to be established in a stepwise procedure eliminating conflicting relationships
and ensuring transitivity of the resulting relation. For a more straightforward identification of pre-
requisites, future investigations should try to select a knowledge domain and/or problems with
unique solution ways.

4. General discussion
The empirical study reported in this paper has put into practice the methodology and demonstrated
the procedure for examining content and application validity of domain ontologies given an exem-
plary concept map representing a small subdomain of geometry. In sum, the results of the present
empirical investigation argue for the validity of the target map from both perspectives, content
validity as well as application validity. The target map can therefore be considered to be validated
and well founded and is ready for use in practice. More importantly and generally, though, the vali-
dation methodology could be meaningfully illustrated and in a concrete validation setting.

The investigation of content validity has been done by comparing the target map with individual
criterion maps that have been empirically collected through a proposition correct-incorrect discrimi-
nation task. Different types of similarity measures have been calculated, in order to compare their
applicability and significance for investigating content validity. Aside from “traditional” similarity
measures (e.g. Goodman & Kruskal, 1979) and measures proposed in the context of concept map
assessment (e.g. Ruiz-Primo et al., 1998) also a non-symmetric similarity approach (Tversky, 1977)
has been inspected. All these measures lead to comparable results and yield fairly high similarity
values for the addressed target map. This, however still leaves open the question whether there is
any best-suited similarity measure that should be applied by default for the purpose of content vali-
dation. For now the usage of several measures is recommended in order to triangulate and cross-
check the different results and to derive a sound conclusive evidence on a target map’s content
validity. In any case, similarity measures from the context of concept map assessment, for calculat-
ing the similarity between student-generated map and an expert concept map (e.g. Ruiz-Primo,
2000), appear highly suitable, since the question motivating these measures can be regarded as
similar, though reciprocal, to the question in case of content validation (i.e. instead of “How similar is
the individual concept map to the expert map?” the question for content validation is “How similar is
the target map to the individual map(s)?”). In case of knowledge assessment the expert concept
map constitutes the criterion for evaluation, in content validation the individual concept maps represent the criteria. The contrast ratio model is suggested for analysis, as it provides the possibility of investigating the similarity between criterion maps and target map in a way to focus the comparison. Actually, this approach provides a framework of realizing different models depending by differently balancing the influence (i.e. weights) of each part of distinctive features. A main question thereby is, of course, how to appropriately set these weights/influences in order to optimally fit the purpose of content validation, in general, or possibly a specific validation objective. It is in fact encouraging that in our study this approach of directional similarity comparison and all other similarity measures applied yield very similar results. More research to gather further conclusions on most suitable and to investigate additionally relevant measures for similarity analysis would be desirable.

In any case, when collecting criterion maps for content validation, but also in case of application validation, it is essential to keep in mind the intended use of a target map. The purpose for which a concept map has been created has an influence on the knowledge covered and depicted, and this needs to be taken into account. Take, for instance, a concept map on photosynthesis. For such kind of domain ontology different purposes can be imagined, which may possibly only differ with respect to the target audience of the map. It will make a difference whether the concept map is intended to be presented to secondary school pupils or to biology students at university. The concept maps for these two teaching purposes will necessarily differ in certain parts, and will in the first case cover rather general and overview knowledge while for the latter case it will present much more detailed knowledge and subject-specific information. Consequently, when posing a concept mapping task for collecting criterion maps, the respective map creators need to be instructed on the intended usage of the map and/or, respectively, selected from an appropriate population. Similarly, the intended field of application of a concept map needs to be carefully considered when choosing a situational performance measure for application validation. In our example, for a concept map with a traditional classroom teaching purpose in mind, problem-solving behaviour has been identified as most appropriate. In contrast to that, for a concept map intended to picture didactics in elementary geometry, for example, actual didactical behaviour from pre-service and more experienced teachers may be suitable.

The validation approach for application validity has been illustrated by representing typical geometry problems of the knowledge domain of interest as substructures of the concept map, and in this way deriving a prerequisite relation in terms of Knowledge Space Theory. The identified prerequisite relationships have given rise to a knowledge structure on the set of problems, which significantly reduces the set of expectable answer patterns down from the number of all on principle possible answer patterns. In this way, the concept map served the derivation and prediction of expected performance behaviour. The theoretically derived knowledge structure has been compared to empirically collected answer patterns, showing that they corresponded well to each other, which is indicative for the validity of the established structure. As the structure is based on the underlying target map to be investigated, the map can be regarded as validated with respect to application validity. One critical issue, though, is the question whether a certain knowledge structure unambiguously results from a specific target map or whether different alternative concept maps may lead to the same knowledge structure. In such case, a validated knowledge structure would give evidence of the validity of both target maps without any conclusion which map is more appropriate. This issue of uniqueness needs to be investigated further in future research.

Overall, future research should be dedicated to broaden the initial encouraging results obtained in this study and to underpin the significance of the proposed methodologies. In particular, it would be interesting to apply the suggested methodological approaches for concept map validation in other, more humanistic knowledge domains. In the humanities actually differing world views and opinions exist to a larger extent than in the more structured domains of natural science, and a larger variance in empirically collected criterion maps for content validation can be expected. Consequently, the significance of the similarity approaches and measures applied for content validation could be investigated further. Furthermore, the application in more humanistic, and therefore, less structured, knowledge domains would also be desirable from the perspective of application validation. As the
approach using Knowledge Space Theory is theoretically applicable for any knowledge domain, this would enable to shed light on our methodology’s potential and significance for validation in knowledge domains of differing degrees of inherent structure. Another interesting aspect for further research would be to investigate concept maps with diverse intended purposes and thus, the use of different behavioural measures as application validation criterion.

5. Conclusion
A concept map representing a domain ontology reflects the perspective or world view of its author(s) and the purpose for which it has been constructed. Hence, there is no “one” correct way of conceptualising and structuring a particular domain (Kennedy et al., 2004). Given a specific purpose, a domain representation is needed that integrates the knowledge of the domain in question. The issue of validating concept maps is critical for their effective use for whatever purpose. This calls for scientifically sound and systematic methods of concept map validation, independent of how a concept map has been created (i.e. manually or automatically). Two different types of validity can be distinguished (Albert & Steiner, 2005a, 2005b)—one referring to the question whether a concept map adequately reflects the content of the respective domain (content validity), and one referring to the issue whether a concept map serves the purpose for that it has been created (application validity). Different kinds of intended application will require different means of validating a target map (Gómez-Pérez, 2001). The presented validation methodology is in line with the categories of ontology evaluation techniques presented by Brank et al. (2005): While application validity can clearly be categorised as an approach “using the ontology in an application and evaluating the results” (p. 166), content validity can be considered an approach “involving comparisons with a source of data ... about the domain to be covered by the ontology” (p. 166). This source of data is in the present case given by individuals’ personal understanding of the domain in question, represented as concept maps. In this way, the presented approach to content validity clearly differs from techniques that are based on the comparison with a gold standard, where there is only one, usually expert-created ontology used as a criterion for validation. The approach also differs from an evaluation by humans who try to directly assess the quality of an ontology with respect to different criteria (Brank et al., 2005).

The present paper outlines an empirical investigation applying the methodological considerations on examining content and application validity, and in this way testing the practicality and significance of the suggested approaches. The applied analysis procedures have proven suitable for the purpose of validation. The different measures used for content validation have all yielded similar results and argue for the content validity of the example target map. The suggested procedure and the performance measure chosen for application validation can be considered appropriate and indicate encouraging results also with respect to the target map’s application validity. All in all, the results of the present investigation are definitely auspicious and are indicative for both, the suitability of the validation methodology, as well as the target map’s content and application validity.

A critical precondition for an effective use of a concept map as a knowledge domain representation is that it reliably covers and depicts the knowledge in question. Therefore, another issue that has to be taken into account actually even before the validity of a target concept map is addressed, is the reliability of the respective map (Albert & Steiner, 2005b), as well as the reliability of the criterion maps collected (Albert et al., 2006). Reliability relates to the consistency and accuracy of a concept map. In this regard, different aspects of reliability can be considered: Consistency over time and consistency over alternative construction methods. Consistency over time refers to the extent to which the generation of a particular concept map is repeatable, i.e. whether the same concept map results when generated again after some time using the same construction method. Another aspect of reliability addresses the consistency over alternative construction methods (e.g. map creation, map completion, etc.) or over different forms of representation (e.g. directed graph, proposition list, etc.). This refers to the extent to which the same concept map representation results, independent of the applied method or format. A reliable concept map represents the same model of knowledge or understanding of a domain regardless of the representation format or time of construction (assuming that
there is no indication of a change in knowledge). For giving evidence of reliability, in both cases (consistency over time and, respectively, alternative construction/representation methods) the extent of similarity between the concept maps is investigated. To this end, similarity measures as proposed for content validation may be utilised.

For collecting concept maps from individuals, representing their personal understanding of the domain of interest, and serving as criterion for examining the target map’s validity, intentionally a rather restrictive method has been used. The correct-incorrect discrimination task has ensured an effective investigation procedure and allowed the collection of data on application validity in the same session. Through the presentation of the target map propositions in combination with distractor items, the vocabulary of the resulting individual domain ontologies has been restricted to a set of predefined concepts and relations. This would be similar, e.g. when using a map completion or map creation task with provided concept and relation sets to be used by participants (compare Steiner et al., 2007). A specific advantage of the method applied in our investigation is seen in the possibility of investigating and comparing competing concept maps implementing different world views within the same task and data collection. The usage of a concept mapping task featuring high “directedness” (Ruiz-Primo, 2000), i.e. large degree of constraints and information provided, necessarily entails a higher expectable similarity between criterion maps and the target map to be validated than when using a concept mapping task with high(er) level of freedom. At the same time, however, with such highly directed (or restricted) methods, it is only possible to a limited extent to get an indication on the “completeness” of a concept map, in terms of whether all the knowledge that is expected to be represented is actually and comprehensively covered (e.g. Gómez-Pérez, 2001; Vrandecic, 2010). The technique used in the present investigation allows avoiding potential problems arising through the usage of different labels for the same concepts and relations, as it would be the case with a low-directedness technique. Through the utilisation of automated matching criteria and algorithms making use of lexical databases such as WordNet (Fellbaum, 1998) or similar resources, it would be possible to take into account synonyms and to identify the similarity between criterion maps and a target map on the basis of the corresponding sets of different lexical realisations or labels (e.g. Brewster, Alani, Dasmahapatra, & Wilks, 2004). Furthermore, the possibility of expressing relations in either the one or the other direction (e.g. “A is part of B” and “B has part A”) needs to be considered in this context. Future work should try to make use of more open methods for the collection of personal ontologies in conjunction with automated procedures and semantic technologies supporting subsequent data analysis.

When talking about content validity, on principle, there are also other ways conceivable for examining this quality aspect than the comparison with empirically gathered criterion maps. A further option for analysing content validity would, for example, be to consult published literature or a collection of documents about the respective knowledge domain, corresponding to the idea of data-driven evaluation through comparison of ontologies with texts (e.g. Brewster et al., 2004). Still another strategy would be to use translations of a concept map into a formal language, such as Common Logic Controlled English (Sowa, 2006) or a formal ontology language (e.g. Corcho & Gómez-Pérez, 2000; Mizoguchi, 2003). These translations could then be presented to the expert(s) who generated the target concept map for judging whether the propositions have been translated correctly and describe what they have been originally intended to express.

The approach presented and demonstrated for application validation is characterised by a systematic investigation of a target map’s ability to predict situational performance. To derive these performance predictions a well-founded theoretical framework, Knowledge Space Theory, is exploited. The procedure applied actually constitutes a method for establishing knowledge structures (Steiner & Albert, 2008) and extends the set of available approaches for structure generation, such as expert queries and mass data analysis (e.g. Koppen & Doignon, 1990; Schrepp, 1999b). A critical assumption, however, in the presented approach is that the association of propositions from the concept map to the problems are reliable and valid. Only if the representation of problems on the concept map is well-founded, the validation procedure can give a sound evidence on the quality of the concept map.
itself. It therefore needs to be ensured that this mapping is done in the most accurate manner possible in order to minimise the risk of introducing any sources of error in deriving the predictions for situational performance (i.e. knowledge states) from the concept map that could, subsequently, lead to erroneous inferences with respect to the map’s validity. Since this issue is similar to the question on the quality of competence assignments to problems in competence-based Knowledge Space Theory, quality criteria and approaches for examining them might be picked up (e.g. Ley & Albert, 2004). If carried out manually, the association between propositions and problems requires an appropriate understanding and expertise of the domain of interest. To enable a more efficient mapping and structuring of problems, an aim of future work is to elaborate on opportunities for exploiting automatic procedures. This could be enabled through the utilisation of process models in order to derive and model the possible solutions ways for problems, and to establish a knowledge structure on them (Schrepp, 1993, 1999a; Witteveen, 1994). Besides, if performance on traditional problems in the context of technology-enhanced learning is selected as a validation criterion, semi-automated approaches making use of problems’ metadata information could support the process of mapping problems to a concept map (for a more detailed discussion please refer to Steiner et al., 2012).

In general, the knowledge structures established and validated in the course of examining the application validity of a concept map can be re-used, e.g. for personalising learning experiences in the context of technology-enhanced learning (Steiner & Albert, 2008). This perspective of reuse would mean shared and reduced costs and supports the potential of the real-world utility of the approach presented. Furthermore, a combined application of concept mapping and Knowledge Space Theory in the context of technology-enhanced learning appears promising. Taking combined advantage of these two methods, on the one hand personalised learning experiences may be realised based on the valid knowledge representation provided through Knowledge Space Theory, while simultaneously visualisations of validated concept map may serve the presentation of learning material or concept mapping as a learning strategy (Oppl, Steiner, & Albert, 2011). Another feasible approach to a combined use of the two methods would be the use of a concept map as navigation interface (e.g. Puntambekar, Stylianou, & Hubscher, 2003) with adaptive navigation support based on Knowledge Space Theory (Krauße & Körndle, 2005). Such an intertwined application of concept maps and Knowledge Space Theory requires not only to have available valid concept map and knowledge structures but also to have a common ground for both approaches of structuring knowledge, which again argues for and confirms the usefulness and benefit of the methodology demonstrated in this paper.

While for the purpose of our study the concept map to be validated has been constructed manually, a target concept map may also have been derived via automated extraction approaches (Chen et al., 2008; Lau et al., 2009; Lee & Segev, 2012). Automated procedures of building concept maps and of ontology learning cannot be assumed to output perfect and sound domain representations and may rather be seen as a starting point or skeleton that need the intervention of human actors and methods for evaluating and validating them (e.g. Zouaq & Nkambou, 2009). The validation methodology presented can contribute to quality assurance of automatically created domain ontologies before their adoption and application for the intended purpose. The validation approaches for content and application validity are also open to automation of certain steps of the evaluation procedure and do not require to be carried out entirely and manually by human users. For content validation the mentioned matching of criterion maps with lexical databases provides, for instance, a starting point. The collection of criterion maps can, of course, be carried out with computer and web support, thus allowing an easy extraction of concepts, relations, and propositions representing personal ontologies for further lexical expansion (synonym matching), as well as for the automatic identification of common and distinct elements of target and criterion maps and the calculation of similarity measures. For application validation the use of formal models of cognitive problem-solving processes in terms of process models and production systems, as well as the exploitation of metadata information, if available, can provide a basis for an effective and semi-automatic identification and representation of situational performance and for the derivation of a structure of predicted behavioural patterns from a concept map. Future work shall continue research in the direction of elaborating the presented methodology towards more automated validity analysis. This is in line with
other recent work on automated ontology evaluation (Brank et al., 2005; Brank, Grobelnik, & Mladenic, 2007; Pammer, 2010; Pammer et al., 2006; Völker, Vrandecic, Sure, & Hotho, 2008), allowing a larger scale applicability of the evaluation methods and more healthy and efficient processing and engineering of knowledge domain representations.

In sum, future research dedicated to the issues and research questions that have been raised and become aware with this initial empirical demonstration appears highly interesting and necessary. By further empirical testing of the proposed validation procedures their contextual conditions of use and applicability can be investigated in more detail and opportunities for their refinement and beneficial application towards optimising evaluation efforts can be identified.

**Funding**
The work presented in this paper was partially supported by the European Community (EC) under the Information Technology (IST) program of the 6th FP for RTD-project iClass [grant number 507922].

**Author details**
Christina M. Steiner1
E-mail: christina.steiner@tugraz.at
Dietrich Albert1,2
E-mail: dietrich.albert@tugraz.at
1 Knowledge Technologies Institute, Graz University of Technology, Inffeldgasse 13/5, Graz 8010, Austria.
2 Institute of Psychology, University of Graz, Universitätsplatz 2/DG, 8010 Graz, Austria.

**Citation information**
Cite this article as: Validating domain ontologies: A methodology exemplified for concept maps. Christina M. Steiner & Dietrich Albert, Cogent Education (2017), 4: 1263006.

**Notes**
1. Please note, that due to unavailability of confidence rating data for the complete set in this case the similarity measures are calculated based on a reduced sample of N = 36.
2. Only solution ways utilising the theorems in right triangles (i.e. Pythagorean Theorem, Altitude Theorem and Euclidean Theorem) are considered, but not solution ways applying trigonometry—as the target concept map is restricted to those theorems. Participants in the investigation have been instructed to work on the problems using the respective theorems on right triangles.

**References**
Albert, D., & Lukas, J. (Eds.). (1999). Knowledge spaces: Theories, empirical research, applications. Mahwah: Lawrence Erlbaum Associates.
Albert, D., & Pilgerstorfer, M. (2007). Preliminary considerations on personalized assessment of logical and moral thinking skills. In D. G. Camhy (Ed.), Philosophical foundations of innovative learning. Proceedings of the International Conference on Philosophy for Children. Conceptus-studien (vol. 18, pp. 55–61). Sankt Augustin: Academia Verlag.
Albert, D., & Steiner, C. M. (2005a). Empirical validation of concept maps: Preliminary methodological considerations. In P. Goodyear, D. G. Sampson, D. J. Yang, Kinshuk, T. Okamoto, R. Hartley, & N. Chen (Eds.), Proceedings of the 5th IEEE International Conference on Advanced Learning Technologies (pp. 216–218). Washington, DC: IEEE Computer Society.
Albert, D., & Steiner, C. M. (2005b). Representing domain knowledge by concept caps: How to validate them? In T. Okamoto, D. Albert, T. Honda, & F. W. Hesse (Eds.), The 2nd Joint Workshop of Cognition and Learning through Media: Communication for Advanced e-Learning (pp. 169–174). Berlin: Japanisch-Deutsches Zentrum.
Albert, D., Hockemeyer, C., & Wesliak, G. (2002). Current trends in eliciting verified personal ontologies. In Proceedings of the UNESCO Conference on Intercultural Education, Jyväskylä.
Albert, D., Steiner, C. M., & Heller, J. (2006, June 9–13). Content validity of concept maps. Paper presented at the Workshop on Cognitive Aspects of Ontology Design and Development (CODE 2006), Second Biennial Conference on Cognitive Science. St. Petersburg.
Almeida, M. B. (2009). A proposal to evaluate ontology content. Applied Ontology, 4, 245–265.
Almeida, M. B., & Barbosa, R. R. (2009). Ontologies in knowledge management support: A case study. Journal of the American Society for Information Science and Technology, 60, 2032–2047.
Anderson, J. R. (1995). Cognitive psychology and its implications (2nd ed.). New York, NY: W.H. Freeman and Company.
Bartz, J. (2005). Statistik für Human- und Sozialwissenschaftler [Statistics for social and human scientists] (6th ed.). Berlin: Springer.
Brank, J., Grobelnik, M., & Mladenic, D. (2005). A survey of ontology evaluation techniques (pp. 166–170). In Proceedings of the Conference on Data Mining and Data Warehouses (SKODD), Ljubljana.
Brank, J., Grobelnik, M., & Mladenic, D. (2007). Automatic evaluation of ontologies. In A. Rao & S. R. Potet (Eds.), Natural language processing and text Mining (pp. 193–219). London: Springer.
http://dx.doi.org/10.1007/978-1-84628-754-1
Brewster, C., Alani, H., Dasmahapatra, S., & Wilks, Y. (2004). Data driven ontology evaluation. In Proceedings of the International Conference on Language Resources and Evaluation (LREC). Lisbon. Retrieved July 2, 2012, from http://eprints.aktors.org/337/02/BrewsterLREC-final.pdf
Chang, K.-E., Sung, Y.-T., Chang, R.-B., & Lin, S.-C. (2005). A new computerized approach based on knowledge space theory. In Proceedings of the Conference on Data Mining and Data Warehouses (SKODD), Ljubljana.
Chen, N. S., Kinshuk, Wei, C.-W., & Chen, H.-J. (2008). Mining e-Learning domain concept map from academic articles. Computers & Education, 50, 1009–1021.
http://dx.doi.org/10.1016/j.compedu.2006.10.001
Cimolino, L., Koy, J., & Miller, A. (2004). Concept mapping for eliciting verified personal ontologies. International Journal of Continuing Engineering Education and Lifelong Learning, 14, 212–228.
http://dx.doi.org/10.1504/IJCEELL.2004.004970
Conlan, O., O’Keeffe, I., Hampson, C., & Heller, J. (2006). Using knowledge space theory to support learner modeling and personalization. In T. Reeves & S. Yamashita (Eds.),
Lau, R. Y. K., Dawei Song, D., Yuefeng Li, Y., Cheung, T. C. H., & Jin-Xing Hao, J.-X. (2009). Toward a fuzzy domain ontology extraction method for adaptive e-Learning. IEEE Transactions on Knowledge and Data Engineering, 21, 800–813. http://dx.doi.org/10.1109/TKDE.2008.137

Leclère, M., Trichtet, F., & Furst, F. (2002). Operationalizing domain ontologies: Towards an ontological level for the SG family. In Foundations and applications of conceptual structure. Contributions to the International Conference on Conceptual Structures (ICCS’02) (pp. 1–14). Bonn: Bürgerschaftliche Akademie der Wissenschaften.

Lee, J. H., & Segev, A. (2012). Knowledge maps for e-learning. Computers & Education, 59, 353–364. http://dx.doi.org/10.1016/j.compedu.2012.01.017

Lewisch, T. (1988). Mathematik. Verstehen – Übern – Anwenden. Band 4. Für die 4. Klasse der Hauptschulen und allgemeinbildunglichen höheren Schulen [Mathematics. Understanding—practicing—applying. Volume 4. For fourth grade secondary schools]. Wien: R. Oldenbourg.

Ley, T., & Albert, D. (2004). Quality criteria for competency assignments: Examples from a project management case study. In P. Dadam & M. Reichert (Eds.), Informatik 2004 – Informatik verbindet. Proceedings of the 34th Annual Meeting of the German Society for Informatics (pp. 342–346). Bonn: Köllen.

Maedche, A., & Staab, S. (2004). Ontology learning. In S. Staab & R. Studer (Eds.), Handbook on ontologies (pp. 173–190). Berlin: Springer.

Miller, N. L., & Cañas, A. J. (2008). A semantic scoring rubric for concept maps: Design and reliability. In A. J. Cañas, P. Reiski, M. Ahlberg, & J. D. Novak (Eds.), Concept mapping: Connecting educators. Proceedings of the Third International Conference on Concept Mapping, Tallinn: OU Vatl Press.

Mizoguchi, R. (2003). Tutorial on ontological engineering: Part 2: Ontology development, tools, and languages. New Generation Computing, 22, 61–96.

Novak, J. D. (1998). Learning, creating and using knowledge: Concept maps as facilitative tools in schools and corporations. Mahwah: Lawrence Erlbaum.

Novak, J. D., & Gowin, D. B. (1984). Learning how to learn. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139173469

Obrst, L., Ceusters, W., Mani, I., Roy, S., & Smith, B. (2007). The evaluation of ontologies. Toward improved semantic interoperability. In C. J. O. Baker and K.-H. Cheung (Eds.), Semantic web: Revolutionizing knowledge discovery in the information sciences (pp. 139–158). New York, NY: Springer.

Opp, S., Steiner, C. M., & Albert, D. (2011). Supporting self-regulated learning with tabletop concept mapping. In M. Mülhäuser, W. Seskin, A. Kaminji, & J. Steimle (Eds.), IAETF Interdisciplinary Approaches to Technology-Enhanced Learning—Interdisziplinäre Zugänge zu technologiegestütztem Lernen (pp. 391–410). Münster: Waxmann.

Pammer, V. (2010). Automatic support for ontology evaluation. Review of entailed statements and assertional effects for OWL ontologies. (Doctoral dissertation). Graz University of Technology, Graz.

Pammer, V., Scheir, P., & Lindstaedt, S. (2006). Ontology coverage check: Support for evaluation in ontology engineering (pp. 123–134). In Proceedings of 2nd Workshop on Formal Ontologies Meet Industry (FOMI), Trento.

Pernao, J., & Aksela, M. (2010). Future chemistry teachers use of knowledge dimensions and high-order cognitive skills in pre-laboratory concept maps. In J. Sánchez, A. J. Cañas, & J. D. Novak (Eds.), Concept maps: Making learning meaningful. Proceedings of the Fourth International Conference on Concept Mapping, Viña de Mar. Retrieved August 7, 2012, from http://cmc.ihmc.us/cmc2010papers/cmc2010-137.pdf

Puntambekar, S., Stylianou, A., & Hubser, R. (2003). Improving navigation and learning in hypertext environments with navigable concept maps. Human-Computer Interaction, 18, 395–428. http://dx.doi.org/10.1207/S15327051HC1804_3

Rappitsch, M., Stamm-Roßnagel, C., & Albert, D. (2004). A performance-space-theoretic approach to the assessment of motor co-ordination in drumming. Unpublished manuscript. University of Graz.

Ruiz-Primo, M. A. (2000). On the use of concept maps as an assessment tool in science. What we have learned so far. Revista Electrónica de Investigación Educativa, 2, 29–53. Retrieved May 30, 2005, from http://redie.uabc.mx/vo2jno1/contents-ruizpri.html

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, 569–600. http://dx.doi.org/10.1002/0022-4898.1996.tb00388x

Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (1998). Comparison of the reliability and validity of scores from two concept-mapping techniques (CSE Technical Report 492). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Schau, C., Mattner, N., Zielik, M., Teague, K. W., & Weber, R. (2001). Select-and-fill-in concept map scores as a measure of students’ connected understanding of science. Educational and Psychological Measurement, 61, 136–158. http://dx.doi.org/10.1177/0013164011791112

Schrepp, M. (1993). Über die Beziehung zwischen kognitiven Prozessen und Wissensräumen beim Problemlösen. [On the relation between cognitive processes and knowledge spaces in problem solving] (Dissertation). University of Heidelberg, Germany.

Schrepp, M. (1999a). An empirical test of a process model for letter series completion problems. In D. Albert & J. Lukas (Eds.), Knowledge spaces: Theories, empirical research, applications (pp. 133–153). Mahwah, NJ: Lawrence Erlbaum.

Schrepp, M. (1999b). Extracting knowledge structures from observed data. British Journal of Mathematical and Statistical Psychology, 52, 213–224. http://dx.doi.org/10.1348/000711099159071

Schrepp, M., Held, T., & Albert, D. (1999). Component-based construction of surmise relations for chess problems. In D. Albert & J. Lukas (Eds.), Knowledge spaces: Theories, empirical research, applications (pp. 41–66). Mahwah: Lawrence Erlbaum Associates.

Šimko, M., & Bielekova, M. (2009). Automatic concept relationships discovery for an adaptive e-course (pp. 171–179). In Proceedings of the 2nd International Conference on Educational Data Mining (EDM), Cordoba.

Simms-Knight, J.E., Upchurch, R. L., Pendergrass, N., Meressi, T., Fortier, P., Tchimev, P., ... Page, M. (2004a). Using concept maps to assess design process knowledge. In Proceedings of the Frontiers in Education Conference [FIE1006/FIE1010]. Savannah, GA: IEEE.

Sowa, J. F. (2006). Conceptual maps. Paper presented at the AERA Annual Meeting, San Francisco, CA. Retrieved April 19, 2006, from http://www.jfsowa.com/talks/cmapping.pdf

Steiner, C. M. (2004). Beziehung zwischen Concept Maps und Kompetenzstrukturen [Interrelation between concept maps and competence structures] (Unpublished diploma thesis). University of Graz, Graz.

Steiner, C. M., & Albert, D. (2008). Personalising learning through prerequisite structures derived from concept maps. In H. Leung, F. Li, R. Lou, & Q. Li (Eds.), Advances in web based learning—ICWL 2007 (vol. 4823, pp. 43–56). Berlin: Springer. http://dx.doi.org/10.1007/978-3-540-78139-4
Appendix A1. Concept map propositions

List of propositions from the target concept map on elementary geometry (theorems in right triangles)

| ID | Concept | Relation | Concept |
|----|---------|----------|---------|
| P1 | right triangle | has side | hypotenuse c |
| P2 | right triangle | has side | cathetus a |
| P3 | right triangle | has side | cathetus b |
| P4 | right triangle | has property | height h |
| P5 | right triangle | has property | right angle |
| P6 | hypotenuse c | is longest side of | right triangle |
| P7 | hypotenuse c | is opposite of | right angle |
| P8 | cathetus a | is adjacent to | right angle |
| P9 | cathetus b | is adjacent to | right angle |
| P10 | height h | passes through | right angle |
| P11 | height h | is perpendicular to | hypotenuse c |
| P12 | height h | cuts into two sections | hypotenuse c |
| P13 | height h | cuts into two right triangles | right triangle |
| P14 | right triangle | has property | area A |
| P15 | area A | is equal to half the product of | cathetus a, cathetus b |
| P16 | area A | is equal to half the product of | hypotenuse c, height h |
| ID | Concept            | Relation          | Concept                                      |
|----|--------------------|-------------------|----------------------------------------------|
| P17| hypotenuse c       | squared           | square of hypotenuse c, (c^2)                |
| P18| cathetus a         | squared           | square of cathetus a-(a^2)                   |
| P19| cathetus b         | squared           | square of cathetus b-(b^2)                   |
| P20| hypotenuse c       | has part          | hypotenuse section p                        |
| P21| hypotenuse c       | has part          | hypotenuse section q                        |
| P22| hypotenuse c       | is equal to the sum of | hypotenuse section p, hypotenuse section q  |
| P23| hypotenuse section p| is adjacent to   | cathetus a                                  |
| P24| hypotenuse section q| is adjacent to   | cathetus b                                  |
| P25| height h           | squared           | square of height h-(h^2)                    |
| P26| right triangle     | can be inscribed in | semicircle                      |
| P27| semicircle         | each triangle inscribed is | right triangle                  |
| P28| semicircle         | has diameter      | hypotenuse c                               |
| P29| square of height h-h^2| is coextensive with rectangle formed by | hypotenuse section p, hypotenuse section q  |
| P30| square of cathetus a-a^2| is coextensive with rectangle formed by | hypotenuse c, hypotenuse section p        |
| P31| square of cathetus b-b^2| is coextensive with rectangle formed by | hypotenuse c, hypotenuse section q        |
| P32| square of hypotenuse c-c^2| is equal to the sum of | square of cathetus a-a^2, square of cathetus b-b^2  |
| P33| circle             | has element       | radius r                                    |
| P34| circle             | has element       | centre M                                    |
| P35| tangent t          | is a              | straight line                                |
| P36| tangent t          | touches           | circle                                       |
| P37| circle, tangent t  | share             | osculation point T                          |
| P38| tangent t          | passes through    | point P outside a circle                   |
| P39| point P outside a circle, centre M | straight line in-between is line PM |                                       |
| P40| osculation point T, centre M | straight line in-between is equals | osculation radius TM |
| P41| osculation radius TM| equals       | radius r                                    |
| P42| point P outside a circle, osculation point T | straight line in-between is | tangent line segment PT |
| P43| tangent line segment PT, osculation radius TM, line PM | form            | right triangle                              |
| P44| osculation radius TM| is perpendicular to | tangent t                                   |
| P45| tangent t          | has property      | tangent line segment PT                     |

**Appendix A2. Distractor items**

List of distractor statements for the correct-incorrect discrimination task

| ID | Concept            | Relation          | Concept                                      |
|----|--------------------|-------------------|----------------------------------------------|
| D1 | right triangle     | has property      | obtuse angle                                 |
| D2 | right triangle     | is a              | triangle with three equal sides              |
| D3 | right triangle     | has property      | volume V                                     |
| D4 | area A             | is equal to the sum of | sides of the triangle                        |
| D5 | area A             | is equal to the product of | cathetus a, cathetus b                       |
| D6 | cathetus a         | is opposite of    | right angle                                  |
| D7 | cathetus b         | longest side of   | right triangle                               |
| ID | Concept                          | Relation            | Concept                          |
|----|---------------------------------|---------------------|---------------------------------|
| D8 | right triangle                  | has side            | secant s                        |
| D9 | hypotenuse c                    | is adjacent to      | right angle                     |
| D10| hypotenuse c                    | is shortest side of | right triangle                  |
| D11| height h                        | cuts into two acute triangles | right triangle                  |
| D12| height h                        | equals              | hypotenuse c                    |
| D13| height h                        | extracted root is   | square of height h-(h²)         |
| D14| hypotenuse c                    | is equal to the product of | hypotenuse section p, hypotenuse section q |
| D15| hypotenuse section p, hypotenuse section q | form               | right angle                     |
| D16| hypotenuse c                    | extracted root is   | square of hypotenuse c-(c²)     |
| D17| hypotenuse c                    | is equal to the sum of | cathetus a, cathetus b           |
| D18| hypotenuse section p            | is perpendicular to | cathetus a                       |
| D19| hypotenuse section q            | is perpendicular to | cathetus b                       |
| D20| hypotenuse section p            | equals              | cathetus a                       |
| D21| hypotenuse section q            | equals              | cathetus b                       |
| D22| height h                        | is parallel to      | hypotenuse c                    |
| D23| semicircle                      | each triangle inscribed is a | isosceles triangle              |
| D24| square of hypotenuse c-(c²)     | is equal to the sum of | cathetus a, cathetus b           |
| D25| square of hypotenuse c-(c²)     | Is coextensive with rectangle formed by | cathetus a, cathetus b           |
| D26| square of hypotenuse c-(c²)     | is coextensive with rectangle formed by | hypotenuse section p, hypotenuse section q |
| D27| square of height h-(h²)         | is coextensive with rectangle formed by | hypotenuse section p, hypotenuse c |
| D28| square of height h-(h²)         | is coextensive with rectangle formed by | hypotenuse section q, hypotenuse c |
| D29| square of cathetus a-(a²)        | is coextensive with rectangle formed by | hypotenuse section p, height h   |
| D30| square of cathetus b-(b²)        | is coextensive with rectangle formed by | hypotenuse section q, height h   |
| D31| tangent t                       | cuts                | circle                          |
| D32| oscillation radius TM           | is parallel to      | tangent t                       |
| D33| circle                          | has element         | height h                        |
| D34| oscillation point T, point P outside a circle | straight line in-between | oscillation radius TM            |
| D35| oscillation radius TM           | equals              | diameter d                      |
| D36| oscillation radius TM           | is perpendicular to | line TM                         |
Appendix A3. Geometry problems

Warming-up exercises:

Problem (1)
Given: \( a = 2 \text{ cm}, \ b = 5 \text{ cm} \)
\( c = 3 \text{ cm}, \ d = 3 \text{ cm} \)
Find: area \( A \)

Problem (2)

Given: trapezium
Transform the given trapezium graphically into a coextensive rectangle!

Problem (3)

Given: \( a = 4 \text{ cm}, \ c = 7 \text{ cm} \)
Find: \( b \)

Problems used for data analysis:

Problem (a)

Given: \( q = 7 \text{ cm}, \ h = 5 \text{ cm} \)
Find: \( p \)
Problem (b)

Given: $q = 7 \text{ cm}$, $p = 6 \text{ cm}$
Find: area $A$

Problem (c)

Given: $b = 5 \text{ cm}$, $q = 3 \text{ cm}$
Find: $p$

Problem (d)

Given: rectangle
Transform the given rectangle graphically into a coextensive square! Only use set square, pair of compasses, pencil without doing any calculations!

Problem (e)

Given: $r = 3 \text{ cm}$, $PM = 7 \text{ cm}$
Find: area $A$

Problem (f)

Given: $a = 5 \text{ cm}$, $c = 8 \text{ cm}$
Find: area $A$
Problem (g)

Given: \( b = 4 \text{ cm}, \ c = 9 \text{ cm} \)
Find: \( h \)

Problem (h)

Given: \( a = 9 \text{ cm}, \ h = 5 \text{ cm} \)
Find: \( \text{area } A \)

Problem (i)

Given: \( b = 3 \text{ cm}, \ c = 9 \text{ cm} \)
Find: \( \text{area } A \)

Problem (j)

Given: rectangle, square

Construct a square that is coextensive with the given rectangle plus the given square.

Only use set square, pair of compasses, pencil without doing any calculations!
## Appendix A4. Assignment of propositions from the target concept map to geometry problems

Each solution way is identified with the corresponding subset of propositions representing the knowledge elements underlying the respective solution approach. The proposition subset $B = \{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}\}$ constitutes basic knowledge on right triangles that is required for each problem and solution way.

### Problems, solution ways and the concept map propositions associated with them

| Problem | Solution way | Associated propositions (in addition to the set $B$) |
|---------|--------------|-----------------------------------------------------|
| a       | $A$          | $\{P_{12}, P_{20}, P_{21}, P_{25}, P_{29}\}$        |
|         | $PK$         | $\{P_{12}, P_{13}, P_{17}, P_{18}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{31}, P_{32}\}$ |
| b       | $HA_1$       | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{20}, P_{21}, P_{25}, P_{29}\}$ |
|         | $HA_2$       | $\{P_{12}, P_{14}, P_{15}, P_{16}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{31}, P_{32}\}$ |
|         | $PKA_1$      | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{31}, P_{32}\}$ |
|         | $PKA_2$      | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{16}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{31}, P_{32}\}$ |
| c       | $K$          | $\{P_{12}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{29}\}$ |
|         | $PH$         | $\{P_{12}, P_{13}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{29}, P_{32}\}$ |
| d       | $HZ$         | $\{P_{12}, P_{18}, P_{21}, P_{25}, P_{26}, P_{27}, P_{28}, P_{29}\}$ |
|         | $K_Z$        | $\{P_{12}, P_{18}, P_{20}, P_{21}, P_{23}, P_{26}, P_{27}, P_{28}, P_{30}\}$ |
| e       | $PKTA(a)$    | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{16}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{31}, P_{32}, P_{33}, P_{34}, P_{35}, P_{36}, P_{37}, P_{38}, P_{39}, P_{40}, P_{41}, P_{42}, P_{43}, P_{44}, P_{45}\}$ |
|         | $PKTA(b)$    | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{16}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{31}, P_{32}, P_{33}, P_{34}, P_{35}, P_{36}, P_{37}, P_{38}, P_{39}, P_{40}, P_{41}, P_{42}, P_{43}, P_{44}, P_{45}\}$ |
|         | $KHTA$       | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{16}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{30}, P_{31}\}$ |
| f       | $PA$         | $\{P_{14}, P_{15}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{29}, P_{31}\}$ |
|         | $PKA$        | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{16}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{30}, P_{32}\}$ |
|         | $KHA$        | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{16}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{30}, P_{32}\}$ |
| g       | $PK$         | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{31}, P_{32}\}$ |
|         | $KH$         | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{31}, P_{32}\}$ |
| h       | $PHA_1$      | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{32}\}$ |
|         | $PHA_2$      | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{32}\}$ |
|         | $PKA_1$      | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{30}, P_{32}\}$ |
|         | $PKA_2$      | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{30}, P_{32}\}$ |
| i       | $PKA$        | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{31}, P_{32}\}$ |
|         | $KHA$        | $\{P_{12}, P_{13}, P_{14}, P_{15}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{31}, P_{32}\}$ |
| j       | $PHZ$        | $\{P_{12}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{30}, P_{32}\}$ |
|         | $PK_1Z$      | $\{P_{12}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{30}, P_{32}\}$ |
|         | $PK_2Z$      | $\{P_{12}, P_{17}, P_{18}, P_{19}, P_{20}, P_{21}, P_{22}, P_{24}, P_{25}, P_{30}, P_{32}\}$ |
Appendix A5. Prerequisites between problems

Documentation of derivation of prerequisite relationships between geometry problems

| Problem | Preliminary prerequisites according to comparison of proposition subsets representing individual solution ways | Adjusted prerequisites after eliminating contradictions | Final prerequisites after adding again missing relationships according to transitivity |
|---------|---------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|----------------------------------------------------------------------------------|
| a       | c                                                                                                            | –                                                      | –                                                                               |
| b       | a, c, f, g, h                                                                                                 | a, g                                                   | a, f, g                                                                          |
| c       | a                                                                                                            | –                                                      | –                                                                               |
| d       | a                                                                                                            | a                                                      | a                                                                               |
| e       | a, b, c, f, g, i                                                                                               | a, b, c, f, g, i                                       | a, b, c, f, g, i                                                                 |
| f       | a, b                                                                                                         | a                                                      | a                                                                               |
| g       | a, c, f                                                                                                       | c, f                                                   | a, c, f                                                                          |
| h       | a, b, f                                                                                                       | a, f                                                   | a, f                                                                            |
| i       | a, b, c, f, g                                                                                                 | a, b, c, f, g                                         | a, b, c, f, g                                                                   |
| j       | a, d                                                                                                         | a, d                                                   | a, d                                                                            |