Modeling of Competences for Students of Engineering Mechanics

F. Musekamp[1], G. Spöttl[1] and M. Mehrafza[1], J.-H. Heine[2], M. Heene[2]
[1] University of Bremen, Bremen, Germany
[2]Ludwig-Maximilians-University Munich, Munich, Germany

Abstract—**Since the implementation of the Bologna process, the quality of engineering study courses should be measured according to what extent academic courses of education facilitate and support the acquisition of engineering competences. So far it is not possible to assess this in a reliable and valid way. The foundation of all competence diagnostics is a model of the competences to be developed as well as the subsequent development and application of test instruments. This article theoretically derives a competence model as the basis for an assessment of learning results in engineering mechanics and places it within the current state of research of empirical educational research. Hypotheses on the competence structure will be derived and, for part of the model, statics, tested on the basis of data from a pilot study.**

Index Terms—**Engineering Science, Mechanical Engineering, Engineering Mechanics, Competence Assessment, Competence framework**

I. INTRODUCTION

One of the reasons why Germany, despite being a land poor in natural resources, enjoys high economic prosperity, is because it has a respectable number of highly-qualified (mechanical) engineers, the academic training of whom therefore takes on particular significance. Even if there has not yet been any diagnosis of an acute lack of engineers up until now, for the medium-term, due to demographic changes as well as ongoing sporadic years of low numbers of graduates in the engineering sciences, a lack of academically-trained graduates can be expected for the future [1]. In light of this future lack there is a disproportionately high amount of drop-outs in mechanical engineering [2]. One of the main causes is considered to be the performance problems, and the accompanying motivational deficits, of drop-outs in the basic engineering courses. Simultaneously, without these subject-specific foundations, the successful acquisition of the necessary interdisciplinary skills must be called into question, since these competences, in particular, must be learned and demonstrated, while dealing with a specific subject matter. In addition, for a career as an engineer, deep and specialized knowledge is indispensible, and the importance of such knowledge tends to be even greater during the first five years of work in such a career.

Before any measures can be taken to improve the training of engineers in Germany, sound diagnostic information must be available which shows whether there are weaknesses in instruction, and if so, where they are to be found. Until now, it has been almost impossible to judge whether the high drop-out rates are the result of a lack in general-education requirements, a lack of motivation or interest, of false assumptions about the requirements for a course of study in the engineering sciences, or of poor teaching. Thus there is a need for reliable instruments for measuring competence within the central specialties of the engineering sciences.

One such sub-section of the engineering sciences is engineering mechanics, or EM. It provides theoretical concepts for application-oriented engineering disciplines (including mechanical engineering) and is a basic discipline for the approximately 370,000 students of the engineering sciences in Germany. It is also an important part of internationally accepted standards of knowledge for the engineering profession [3].

Within the framework of the study presented here, two test instruments for the measurement of mechanical-engineering competences will be developed and validated by the end of 2013, in order to help generate diagnostic information about students' learning progress.

The starting point and the most important basis for the assessment of general competence as well as for competences specific to engineering mechanics (EM) is a competence model (which we shall refer to as the EM Model'). Its design is of primary importance for the explanatory power of test instruments which will be based on it as well as the generalization of the results achieved. The goal of this article is thus to position the EM Model within the current landscape of empirical research on education and training, to theoretically ground the specific design of the EM Model and to find first evidence for its empirical validity in the area of statics on the basis of pilot data.

Prior to this, we will sketch out the aims of the study 'Modeling and Measuring Competences of Engineering Mechanics within the Training of Mechanical Engineers (KOM-ING)' as well as the underlying theoretical framework of this conception of competence.

II. AIM OF THE STUDY

The aim of the study is to validate a competence model for EM as well as the construction and testing of two related instruments for measurement. We will examine whether the EM Model can be empirically proven, and (1) to what extent the EM competence of students can be differentiated and described on the basis of the dimensions of competence being postulated. (2) If these questions can be answered in the affirmative, then the first instrument makes it possible to account for different training pathways, and can be used for comparison work at the institutional level (summative assessment). The second instrument then delivers more profound information, for the sub discipline static engineering of EM, about the desired and

actual states of the learning process for particular points in time, which can be used for the configuration of teaching. (formative assessment)

III. THEORETICAL CONSIDERATIONS AND STATE OF THE RESEARCH

A. *On the Concept of Competence*

Competence is a "theory-relative" concept (as discussed in [4]). Competence modeling thus needs to be located within the diversity of different views of competence. The starting point for these considerations is the constitutive characteristic of the concept of competence, which applies to the ability to use one's mental skills in new situations [5]. Situations are, in turn, characterized through conditions of the surrounding environment, which are presented to the actor as objective circumstances (often described as 'context'). Reference [6] make it clear that context cannot (or not only) be interpreted as a naturally-existing condition but, in most situations, is determined by a socially-assigned responsibility. They conceptualize competence as a 'product of interdependence' between 'social responsibility' and 'individual mental variables' [6]. Thus, competence is a dual construct, through which people are seen as being competent who can accomplish what they are supposed to accomplish [7]. Societal responsibilities are always closely connected to expectations of behavior, which can vary in their intensity. Against the background of these conditions and restrictions, the EM Model then has to explicitly differentiate between objective requirements (referred to below as 'the context side') and mental disposition in order to determine which are crucial to meeting the requirements (referred to below as 'the disposition side').

Thus, the EM Model is oriented towards the German Research Foundation's (DFG) Priority Program 'Competence Models for Assessing Individual Learning Outcomes and Evaluating Educational Processes' [8], and its construct of competence, excluding volitional and motivational aspects. In that program competence is defined as 'domain-specific cognitive dispositions that are required to successfully cope with certain situations or tasks, and that are acquired by learning processes' ([8], p. 68). In order to describe the 'disposition side,' generally psychometric models of competence, which can be differentiated into structure and level models [9] are used. Structure models describe quantitatively different mental characteristics (so-called constructs or dimensions), which can be differentiated on the basis of factor-analytical methods (for example subject specific knowledge). Level models provide information about how the content of high and low manifestations can be described, but do not necessarily give predictions for empirical competence development [10].

B. *State of the Research*

In order to define the normative requirements and to derive a hypothetical construct of the cognitive disposition required for the mastery of these norms, we will discuss the state of the research in academic research on education as well as scientific research on competence in general-education schools. For the context side of competence we will rely on the example of the European Qualification Framework (EQF, [11]) as a relevant framework for qualification. Empirical requirements analyses of working engineers (for example [12]) will not be considered, since these have much too low a level of detail to be used with a focus on the basic subjects of EM. Political demands from engineering associations, which primarily emphasize the importance of specialized competence over generic competence, (for example, [13]) will also not be discussed here due to a lack of space.

1) *Context: Behavioral Expectations and Responsibilities*

Below we will equate competence in the sense of responsibility, with 'socially-framed norms of behavior' because the allocation of responsibilities is always connected to the expectation of their fulfillment. Responsibilities are therefore normative stipulations which arise from negotiations between social or political stakeholders. The results of such negotiations can be empirically described and used for the formulation of the context side of a competence model.

In mechanical engineering, someone is 'responsible' if they have a qualification in the form of a higher-education degree as a (mechanical) engineer. Although qualification frameworks like the EQF primarily serve as an instrument for transnational comparability of graduation levels, they are also seen, particularly in connection with competence orientation, as 'institutional requirements of overriding importance' [14]. This is easy to comprehend with a glance at the structure of the European Qualifications Framework, or other frameworks, where the argumentation likewise preserves its plausibility.

The EQF consists of eight levels, of which three are academic courses of study (6 Bachelor, 7 Master, 8 Doctorate). These levels are described with the help of mental characteristics as 'knowledge', 'skills', and 'competence', meaning 'the proven ability to use knowledge, skills, as well as personal, social, and methodological abilities in work or learning situations for professional or personal development.' [11] The categories for the qualification framework are more closely defined with the help of descriptors, which formulate personal learning outcomes. The assigning to levels of the category 'competence' will be accomplished through situational conditions and contextual characteristics, as made clear in the following formulations:

„Leading of complex specialized or professional activities or projects and the taking over of responsibility *in unpredictable work or learning situations*" (Level 6)

„Leading and organizing complex and unpredictable *work or learning contexts* which require new strategic approaches" (Level 7, [11], p. 13; italics the authors).

The use of context characteristics in the descriptors shows that requirements will be different according to context, although the increase is not coherently formulated in every case. For the individual, this means that a certain *behavior* will be *expected* for the particular context of the level in question.

Having said that, within the framework of the EM Model *socially agreed-upon behavioral expectations and/or context characteristics* in the sense of requirements should be formulated. (a). It is also clear, however, that standards of accomplishment must be made more detailed and more stringent than, for example, is the case in standards for qualifications. (b)

2)   Mental Disposition: Knowledge and Cognitive Processes

In order to achieve an empirically visible and profitable structure for the disposition side of the EM Model, the current state of research will be examined with regards to which psychological or subject-oriented aspects are empirically meaningful for the structural organization and level descriptions of EM competences.

Neither national nor international psychometrically-validated competence models for engineers in general or for EM in particular exist. [15] There has been a recent attempt at modeling and psychometrically testing engineering competence. The 'Tertiary Engineering Capability Assessment' (TECA, [16]) was developed within the framework of the OECD Program 'Assessment Of Higher Education Learning Outcomes' (Ahelo, [17]). This program seeks to be able to compare universities' output internationally (summative Assessment). The TECA competence model contains, in addition to generic competences ('Professional Attributes') also 'discipline-related competences' ('Technical Knowledge' and 'Engineering Process'). Empirical validation is intended to be achieved through computer-based tests and probability-based test models, however, these still remain to be developed. Since EM competences only play an inferior role within the TECA model and there are no further academic competence models for engineering, general educational models from the natural sciences will be used.

Almost all the recently-developed competence models for physics and mathematics claiming to be psychometrically testable, cover at least the following two dimensions: 'cognitive activities,' which can be understood as being analogous to the *cognitive processes'* discussed by [18], and *'content areas.'* Thus, [19] differentiate, for example, between 'cognitive processes' (p. 168) and 'basic concepts/content areas' (p 171), and [20] between 'cognitive activities' and 'central ideas'. None of the approaches explicitly differentiate whether they deal with content areas which fall under the 'Structure of Discipline' [21] or their internal cognitive representation in the sense of 'knowledge' (see for example [18]) or a congruent structure of both aspects. In addition, dimensions of competence models of the natural sciences often are not attributable to the disposition side of competence, but rather describe context-characteristics of achievement such as 'length of text' or 'complexity of problem solving' [20]. It should be taken into account that 'cognitive processes' do not deal with individual thought processes connected to particular situations, as the term suggests, but instead with a more long-term disposition towards bringing such processes forth.

The internal structures of *cognitive processes* are diverse in their number and quality: Reference [19] differentiate, for example, seven 'cognitive processes' such as 'verbalizing complex circumstances', 'dealing with numbers' or 'convergent thinking[1]' (ibid. p. 177). While here the cognitive processes are logically independent of one another, the construct of 'mathematical literacy' within Pisa [22] refers back to the processes of Mathematical modelling following [23] and has a chronologically ordered structure. Day-to-day circumstances must be trans-

lated into a mathematical model and be inner mathematically processed in order to then be able to be interpreted. Finally, the results are translated back to the day-to-day situation and validated. In contrast to [19] 'mathematical literacy' is not dimensioned separately along the four partial processes of Mathematical modeling. Instead, it is assumed that in day-to-day mathematical tasks, all four processes must occur, so that these are built into all PISA test tasks with changing amounts of emphasis.

The *content dimensions* of the competence models being currently discussed naturally differ along the diversity of the domains being examined. However, generally models whose subject-area structure is connected to particular disciplines can be differentiated from those which have non-disciplinary concepts as a content dimension. Prototypically, the modeling within the framework of the PISA studies is led by 'big ideas' ([22], p. 12). This is understood to mean „mathematical concepts which are strongly connected to one another but can be seen as being grouped under a common aspect" [22]. In contrast to this, for example, within the national supplementary tests of PISA 2000, a difference is made between the subject-analytical areas of arithmetic, algebra, geometry, and stochastic [24] in order to improve the informational content of the results from a national perspective.

The acceptance of a hierarchical structure of cognitive processes is increasingly being rejected in favor of categorical structures. Instead, levels are supposed and in part empirically proven based on content areas or content concepts (and their use). Reference [20] for example, sees competence development in physics (secondary education) as the expansion and differentiation of knowledge, which accompany an increasing conceptual understanding. (compare p. 95) They test the hypothesis that pupils' concept of energy evolves hierarchically in ascending order along the four *content*-specific levels: forms and sources of energy, (1), changes and transfer of energy, (2), loss of energy, (3), and preservation of energy (4). Conforming to expectations, it turns out that there is a significant statistical connection between the empirical difficulty of a task and the stage of development.

At the moment, instruments for compiling skills are also being developed internationally in the academic world, with a focus on engineering mechanics. In order to do so, so-called 'concept inventories' [25][26], are being developed which raise the claim of being able to pinpoint conceptual understanding as well as misunderstanding. The 'concept inventories' are partially pyschometrically validated [27][28] and anticipate direct use for the development of teaching curricula, because the hope is that inappropriate conceptual knowledge can be diagnosed and, based on this diagnosis, changed through teaching. [29][30]. However, the question as to whether this expectation can really be fulfilled is not undisputed, as was shown in an independant test of the 'force concept inventory' [31]. The validity of this instrument, which originally claimed to be able to differentiate between test subjects who had a Newtonian conception of force from test subjects with various less-informed conceptions of force, is relativized through [32] as well as [33]. On the basis of different data, they arrive at the conclusion that the FCI cannot reliably differentiate between the six dimensions (kinematics, principle of inertia, force and momentum, interaction of forces, overlapping of forces, special forces)

---

[1]   Convergent thinking means to reason correctly from given information.

and therefore, cannot make a diagnosis of pupils' conceptions of the concept of force possible.

In view of the present state of the research, the following four theses for the explanation of the disposition side of a EM Model become clear: context-specific achievement dispositions are only conceivable as a combination of cognitive processes and knowledge contents. (a) In order to make a detailed diagnosis of students' skills, it makes sense to understand cognitive processes as individual dimensions, while keeping in mind that during the solving of real-life EM tasks, several processes are often taking place at the same time. (b) The highly disciplined structure of EM and the academic teaching involved calls for a subject-specific structure of its knowledge contents, for example, structuring based on statics, dynamics, and so on. (c) A categorical structuring of cognitive processes is preferred to a hierarchical ordering, although the development of competence along the content areas that are based on one another should not be excluded. (d)

C. Competence Model for Mechanical Engineering

In light of the state of the research the EM Model will be constructed in the following way: in order to do justice to the dual structure of competence, the EM Model contains two matrices. (see Fig. 1) The matrix on the left defines the objective requirements of competence in the sense of supra-individually valid expectations of behavior (context). The right matrix represents the internal, mental requirements which are necessary for the mastery of the context side (context-specific achievement dispositions). While the context side of the EM Model represents the setting of norms, the disposition side can be seen as a hypothetical construct which requires empirical testing.

The context and disposition sides are related to each other insofar as the solution of every requirement that is defined on the context side requires a specific combination of the dispositions. In order to handle a real statics task with a given level of difficulty, the four cognitive processes must take place with the accessible knowledge in dealing with rigid bodies in equilibrium.

1) Context

The requirements within the area of EM can, generally speaking, result from the perspective of the subject to be learned, from the perspective of the academic disciplines which are based on it, such as design methods, or from the world of work. The context side of the EM Model concentrates on the question as to what is expected from *students* of mechanical engineering, that is, under what conditions they are supposed to achieve EM-related output.

In basic mechanical engineering courses, it is expected that students can intellectually comprehend and use the specialized theories and concepts of EM as well as the methods of thinking and working which are specific to the subject. The application of subject-specific basic workplace knowledge is, in contrast, *not* a part of learning expectations (although naturally, these basics do get used in professional tasks). Usually, the tasks are accomplished with simple tools such as paper, pencil, and pocket calculator, whereas the use of tools that are relevant to the working world (such as the computer) is not expected.

In light of this, context is defined by content from the commonly taught EM **Subject Matter**, **(I)** and the **level of Requirements (A) which are determined by the** so-called boundary conditions of mechanical objects.

For **(I)**, First of all, there is a documented content structure that exists in textbooks. Most of the multi-volume textbooks deal with statics, mechanics of materials, and dynamics, in this order. Further areas such as fluid mechanics will not be dealt with in the following, since they are taught to a different level depending on the institution.

For **(A)**: requirements vary with the complexity of the mechanical objects being dealt with and it is supposed that the level of requirements can be depicted by the number and type of the so-called boundary conditions, which are statements concerning the displacement and stress of the boundaries of objects in mechanical systems [34]. The number and type of boundary conditions can be objectively identified for EM tasks and can be limited to an area where the solving of tasks is possible with paper, pencil, and calculator. It must be mentioned here that the three-fold classification of the requirements axis is not a part of the theory but merely an arbitrary placeholder for a scale that will probably be continuous when it is empirically tested.

In summary, it is expected of students of engineering sciences within the sub discipline of EM that they solve tasks which are to a large extent clearly assigned to one of the three subject areas of statics, mechanics of materials, or dynamics, (1) and in which only a few or simplified boundary conditions are to be met, (2) so that they can be solved with the help of paper, pencil, and calculator. (3) Both dimensions for the description of competence within the aspect of responsibility (context) are thus 'subject matter' and 'level of requirements.' The mental dispositions for achievement are described below with regards to this context.

2) Context-Specific Perfomance Disposition

The cognitive processes which underlie the solving of EM tasks are very similar to the mathematical modeling described by [23]. In EM, real existing circumstances must also be transferred into models and their validity must be tested: (prospective) engineers have to analyze problems in EM, understand the fundamentals, and transfer a real object into a physical model. The mathematical problems which follow must be solved, and the interpretation of the results, which connects back to the real object, closes the circle [35]. Since diagnostic information should serve not only for achievement comparison at the institutional level, but also for the improvement of instruction (formative assessment), in contrast to the PISA approach, a separate recording of these processes in the sense of their individual dimensions is appropriate.

EM-specific content areas on the disposition side of the model can only be interpreted in the sense of individually-accessible knowledge. In general, knowledge is defined as 'permanent availabilty of information understood' [36] and can be further differentiated into knowledge about concepts as well as knowledge about methods. In EM, knowledge of concepts (K) and knowledge of methods (V) are largely very specifically connected (K/V) but can be conceptually subsumed under the external, previously-existing content structure of EM, 'statics', 'mechanics of materials.' and 'dynamics.' For example, 'knowledge about the concept of support reaction and the methods for defining it' belongs to statics. Similarly 'defining the bending line' is an example that falls under mechanics of materials and 'defining equation of motion' is an example for dynamics. In order to conceptually differentiate be-

tween static, mechanics of materials, and dynamic at the external level from the internal level, knowledge of EM concepts and methods (K/V) are different in 'rigid bodies in equilibrium,' 'elastic bodies in equilibrium,' and 'moving bodies.' The performance dispositions arise from the processes of **mechanical modeling (P)** which are used in EM, with the help of individual knowledge about **concepts and methods of EM (K/V).** (see Fig. 1).

It must be kept in mind that the four cognitive processes of mechanical analysis are handled with differing intensity in typical EM courses and textbooks. For example, in mechanical engineering, the first step of abstraction is usually omitted and one starts directly with mechanical models. In other science courses of study within the engineering sciences, it is very possible that there are additional differences in which subject matter is weighted. It can be assumed, however, that teaching (and hopefully learning) the entire process of mechanical analysis would facilitate the transfer of EM competence into more specialized, advanced courses as well as the working world.

D.   Learning and Teaching-Theoretical Classification and Hypothesis

The choice of psychologically-characterized terms used up until now should not disguise the fact that the EM Model is a modern description of teaching goals, as formulated within empirical research on education and training. Within that research, a teaching aim is „seen as a personality trait which in turn is defined through a set of tasks" [37]. The disposition side of the EM Model, in this sense, specifies what should be done with which content (content x process). The sets of tasks which thus result are labeled with terms that represent personality traits (for example knowledge). The disposition side of the model is to be understood as a teaching goal of academic instruction within the subject of EM. The personality traits result from the intersection between process and content dimensions and can be interpreted as EM-specific skills. Therefore the left lower cell of the disposition side is to be seen as the (partial) ability to abstract real bodies in equilibrium onto a mechanical model.

The advantages of the EM Model lie in the precise formulation of the theoretical term competence (as a teaching goal) which allows for strict testing.

Here it is assumed that the response behavior in each of the four proclaimed process dimensions within the area of statics each lead back to a homogeneous construct of skills (Hypotheses 1: Validity of the Rasch Model) and that these four skills are independent of one another (Hypothesis 2: Dimensionality of the EM Model). It should be mentioned that the data for the dimensions mechanics of materials and dynamics are not available at the time of writing.

IV.   RESEARCH DESIGN AND METHODS

A.   Survey instrument

In order to validate the EM Model, a sample of content-valid examination tasks that is as representative as possible is taken from each of the four statics cells of the right-hand matrix (see Fig. 1) The survey instrument thus varies the statics content in the form of 120 items across all four process dimensions of the EM Model. Based on the typical requirements in EM courses the tasks are preponderantly open ones. Because of the large number of items needing to be tested, all 120 statics items are arranged in 15 test booklets in a Balanced Incomplete Block Design (BIBD, see, for example, [39]) which each contain 32 items. The four different dimensions are equally represented within each test paper with eight items each.

B.   Data

The results presented below are based on cross-sectional data collection from pilot studies, conducted at the end of the first departmental semester in the Winter Semester 2012/13 at two universities (n=258) and a technical college (n=20). Since the sample yield was markedly lower than expected, the total sample consists of 278 engineering students of mechanical engineering. Because of the BIBD and individual missings, every test item was attempted 19 to 77 times.
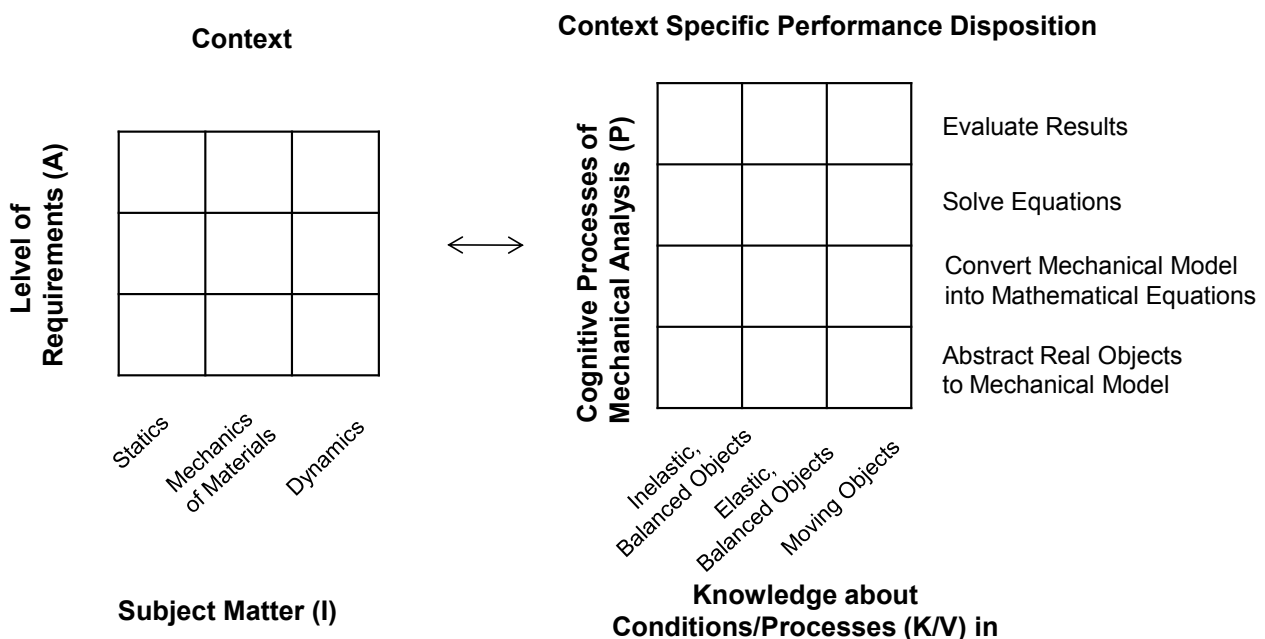

Figure 1.   Competence Model for Engineering Mechanics (EM Model)

C.   Methods

1)   Rater Reliability

The foundation for a validation analysis is, firstly, the highest degree of analytical objectivity possible; in other words, there should be a high degree of agreement between the individuals responsible for the correcting of the tests. The tasks in the test were corrected by a total of nine responsible individuals. For this purpose, a correction scheme that had been checked by EM experts was used. This scheme contained 2 to 4 categories depending on the task in question, in order to represent the 'correctness' of the solutions in an ordinal sequence. Approximately a third of the test papers worked on were independently and doubly evaluated. In the selection of papers which were to be corrected twice, each of the 15 different test papers should be present with equal probability. (p= 1/15) At the same time, the papers selected should represent the lowest as well as the highest skill range of the participants. The random drawing that is stratified in such a way (by test paper and according to skill level) is realized by:

- The number of test tasks which were answered entirely correctly was determined, on the basis of the first ratings, as a first approximation of the respective levels of skill of the students, and the sampling was divided into two halves (median split) .
- From the two sub-samples which resulted, three were randomly chosen for each of the 15 different test papers. (n= 2x3 x 15 = 90 papers)

On the basis of this data, the ICC1 for the guesser conformity and the ICC1k for the reliability of the overall assessment are calculated (see [47]).

2)   Scaling

In order to validate the competence model, the data are scaled according to the psychometrical requirements of the Rasch Model [40] the Partial Credit Model (PCM) [41] respectively. This is usually carried out by means of Maximum Likelihood Estimation Methods (for example, as in [42]). If only few cases are available, and where countless values are missing, the identification of the maximum of the Likelihood Function can be problematic, and the exactness of calculation can be reduced, under certain conditions [43]. This was the case with the pilot study data used here, because of the BIBD as well as unexpectedly low participant numbers. In addition, the test papers, with their 32 questions, proved to be too extensive, so that on average only two thirds of the tasks were attempted. Thus, the multiple-level answer format was recoded into a dichotomous format (true/false) and scaled with help of explicit parameter calculation following the principle of pairwise comparison of contingent solution probabilities. (see [44]). Here, the frequency of solution of item j, under the condition of item i not being solved, is calculated for all items. After further reconfiguration of the symmetrical matrix resulting from this, the item parameter of the dichotomous Rasch Model can be calculated [45]. This approach proved to be less susceptible to high percentages of missing data when compared to the Maximum Likelihood Method [46]. The scaling is carried out with the use of R Software and the package „pairwise" [47].

3)   Model Validity

As an initial, still-to-be-deepened approach to testing the validity of the Rasch Model, a graphical model test was carried out. The scaling of the data was done independently on two sub-samples, testing if the resulting item difficulty corresponds in both samples. This criterion for model validity, known as person homogeneity, is a requirement of the Rasch Model. The total sample has been randomly divided into two subsamples.

In order to preliminary test the dimensionality of the EM Model in the content area of statics, person ability parameter were determined as weighted likelihood estimates (wle) and were correlated according to the four cognitive process dimensions. The scaling was realised using the pairwise estimates for item parameters.

V.   PROVISIONAL RESULTS FOR MODEL VALIDITY

A.   Rater Reliability

Since, through BIBD, every single test item appeared four times in each test booklet (replication factor r=4), each of the 120 items could have a maximum of 24 double-valued solutions (3 random samplings x 2 competence levels x4 repetitions). Actually, the number of cases per item ranged from n=0 to n = 22 due to individual missings. Across all four dimensions the values ranged from r_{ICCI} = -.23 in 'Abstract' to r_{ICCI} = 1 in the 'Convert', 'Solve' and 'Evaluate' dimension. The values for the reliability of the overall assessment ranged from r_{ICCIk} =-.6 to r_{ICCIk} = 1. (An overview of the coefficients per task can be found in „Supplementary File A".) These parameters indicate severe violations of rating objectivity and consequently for the instruments reliability. For this reason the following calculations of model validity were carried out with a reduced set of items; items with an ICCI of <0.5 have been excluded. Depending on the cognitive process there remained between 16 ('Abstract') and 20 items (Convert/Evaluate) for further analysis. In addition, the 'Solve' dimension needed to be reduced, because four items were hardly not attempted by the students.

B.   Hypothesis 1: Model Validity

Fig. 2 shows the results of the item parameter calculation for the dimension 'Abstract' following the „pairwise method" (16 items). The x-axis represents the parameter that result from scaling of random subsample 1 and the y-axis shows the parameter based on random subsample two. With good model fit, it would be expected that the entire sequence of item parameters would not differ, and that all points lie on the bisecting line.

While there are no outliers detectable for the items of the 'Abstract' and the 'Evaluate' dimensions, (see Fig. 2 for the 'Abstract' dimension), for the 'Convert' dimension and for the 'Solve' dimension, one and three items deviate from the bisecting line, respectively, and these are outside the 95 % confidence ellipse (see „Supplementary File B").
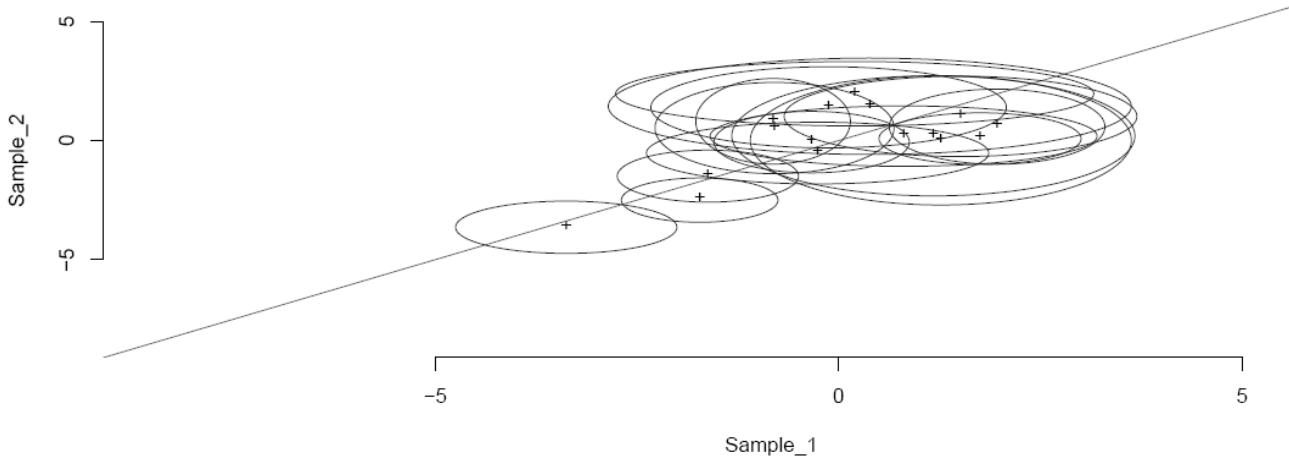
Figure 2.   Graphical Model Check for the Statics Abstract dimension

This small amount of outliers can very carefully be interpreted as a hint for the validity of the Rasch Model at least within the dimensions 'Abstract', 'Convert', and 'Evaluate'. The 'Solve' dimension, however, which deals with the necessary mathematical capabilities to generate numerical results for mechanical problems, shows serious weaknesses and six further items needed to be excluded because of bad fit to the Rasch model.

After deleting all items showing graphical misfit, between 12 ('Solve') and 20 items ('Evaluate') still remain for determing the person ability parameters on each dimension.

C. Hypothesis 2: Dimensionality

Since the BIBD was not balanced for equal booklet difficulty, the booklets have been checked for systematic bias in mean difficulty before person wle estimates were correlated by each of the four dimensions of mechanical analysis (see TABLE I).

TABLE I.
RANGE OF MEAN DIFFICULTIES OF THE 15 TEST BOOKLETS

	Abstract	Convert	Solve	Evaluate
Mean Booklet Sigma (min)	-1.43	-1.29	-2.03	-.87
Mean Booklet Sigma (max)	1.53	.81	2.31	.84

The 'Solve' dimension in particular, suffered considerably from the variation of mean difficulties between the 15 test booklets. Therefore, the wle estimates were corrected according to the mean difficulty level of the booklets by adding the booklet difficulty parameter to the individual wle estimate.

The correlations between the four dimensions 'Abstract', 'Convert', 'Solve', and 'Evaluate' reach values between $r=-.16$ and $r=.44$ based on uncorrected wle-estimates (see TABLE II, below the diagonal). The correlation of the wle estimates when the values were corrected for test booklets difficulty varied between $r=-.28$ and $r=.33$ (see TABLE II, above the diagonal).

While correlations between the dimensions 'Abstract' and 'Convert' as well as between 'Solve' and 'Evaluate' come to comparable results independently of the method of approach, considerable differences appear in the corre-

lations between corrected and uncorrected values. Especially, the relations between 'Abstract' and 'Solve' or 'Abstract' and 'Evaluate' are not clear without ambiguity, due to large absolute differences in correlation coefficients.

TABLE II.
CORRELATION MATRIX FOR COGNITIVE PROCESS DIMENSIONS
(CORRECTED ABOVE AND UNCORRECTED BELOW THE DIAGONAL)

	Abstract	Convert	Solve	Evaluate
Abstract		.16	.23	-.28
Convert	.13		.33	-.05
Solve	.44	.39		-.01
Evaluate	-.16	.07	-.04	

On the one hand these differences underline the uncertainty that can still be found in the sparse pilot data. On the other hand, none of the correlations exceed an absolute value of .44 which can still be deemed as relatively low, compared to other studies in educational research ([50][51]). Thus, we would interpret this result as evidence that EM Competence in the content area of statics comprises discrete process dimensions rather than an overall statics capability. Nevertheless, this interpretation is a preliminary one, since it is likely that all correlations are underestimated, due to the unsatisfactory rater reliability up until this point.

VI. SUMMARY AND PERSPECTIVE

This article intended to present a competence model for the subject matter of Engineering Mechanics and preliminary results of its empirical validation on the basis of pilot data.

The developed measurement instrument shall deliver information on "student responses […that] can be used to shape and improve the student's competence" ([52], p. 120). According to the EM Model, this information refers to four sub capabilities that are supposed to be important not only for the accomplishment of requirements within academic courses, but also partially, also for meeting the demands of the working environment. The model shall reveal which subject matter should be taught and which competences should be developed. The test items howev-

er, indicate what the students already have a command of, what they still need to learn, and within which fields.

The presented instrument for measuring statics competences has not yet reached its applicability for psychometrical reasons. Rater reliability in particular, turned out to be lower than expected and necessary. For that reason it is forseen to train the raters more intensively prior to the main study in January/February 2014 using wrong and partially right answers given by the students during the study. This has not been possible to implement for the analysis presented here, due to the tight time schedule. Furthermore, improvements of the correction scheme are supposed to lead to better reliability coefficients. Moreover, during the pilot study the insufficient test time could been identified as a probable cause for the, generally too high, item difficulty levels. It is therefore predictable, that the validity of the model could be improved through an abridgement of the test booklets.

From a theoretical perspective, however, the presented analysis shows promising results. The competence model assumes a fourfold dimensional structure. These dimensions could be supported empirically and turned out to correlate with each other to a moderate extent. This supports the validity of the EM Model, indicating that EM competence in the content area of statics consists of four distinct cognitive processes of mechanical analysis, rather than of a unified capability in the content area of statics.

Another piece of evidence suppurting the validity of the EM Model delivered [53]: They developed learning sets in the sense of [54] which describe a hierarchical structure of sub capabilities necessary to solve the items which are assigned to each statics cell of the model. Reference [53] demonstrate on the basis of the same data set that has been used to test the dimensionality in this article that the empirical item difficulty parameters depend significantly on the extension of these theortically claimed learning sets.

Nevertheless, it has to be mentioned, that these results need to be interpreted with caution for at least two more reasons: Firstly, the test booklets could only be administered to a very small sample of EM students, leading to a sparse data set. Secondly, model deviations can be detected especially in the 'Solve' dimension and cannot be solely explained by the high level of missing data or insufficient rating objectivity. In general, 'Solve' items turned out to be too difficult and it was this dimension in particular, which showed big differences in the mean booklet difficulties and consequently in the corresponding corrected respectively uncorrected wle estimates of person abilities. Aside from this booklet effect, it has to be assumed that item construction errors as misleading formulations or erroneous formulars contributed to this result.

As a short summary of the experiences from this pilot study it seems feasible to deal with the current psychometrical shortcomings of the instrument. From a theoretical perspective, there is a strong possibility that the measurement instrument will enable assessment based teaching in the future.

REFERENCES

[1] W. Biersack, A. Kettner, and F. Schreyer, Fachkräftebedarf: Engpässe, aber noch kein allgemeiner Ingenieurmangel. IAB-Kurzbericht: 16, 2007.

[2] U. Heublein, C. Hutzsch, J. Schreiber, D. Sommer, and G. Besuch, Ursachen des Studienabbruchsin Bachelor- und inherkömmlichen Studiengängen: Ergebnisse einer bundesweiten Befragung von Exmatrikulierten des Studienjahres 2007/08 Hannover: HIS Hochschul-Informations-System GmbH, 2009.

[3] C. Ferguson, Defining the Australian mechanical engineer. European Journal of Engineering Education, 31 (4), pp. 471-485, 2006. http://dx.doi.org/10.1080/03043790600676646

[4] J. Erpenbeck and L. von Rosenstiel, Einführung: In J. Erpenbeck & L. von Rosenstiel (Hrsg.), Handbuch Kompetenzmessung : erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis. Stuttgart: Schäffer-Poeschel, 2007.

[5] E. Klieme and J. Hartig, Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), Kompetenzdiagnostik. Zeitschrift für Erziehungswissenschaft. Sonderheft 8 (S. 11-29). Wiesbaden: VS Verlag für Sozialwissenschaften, 2007.

[6] G. A. Straka and G. Macke, Berufliche Kompetenz: Handeln können, wollen und dürfen: Zur Klärung eines diffusen Begriffs. BWP (3), pp.14-17, 2009.

[7] G. Spöttl and F. Musekamp, Berufsstrukturen und Messen beruflicher Kompetenz. Berufsbildung, 63 (118), 2009, pp. 20-23.

[8] K. Koeppen, J. Hartig, E. Klieme, and D. Leutner, Current Issues in Competence Modeling and Assessment. In: *Zeitschrift für Psychologie / Journal of Psychology* 216 (2), pp. 61–73, 2008.

[9] J. Hartig and E. Klieme. Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), Leistung- und Leistungsdiagnostik. Berlin: Springer, 2006, pp. 127-143. http://dx.doi.org/10.1007/3-540-33020-8_9

[10] H. Schecker and I. Parchmann. Modellierung naturwissenschaftlicher Kompetenz. ZfDN (12), pp. 45-66, 2006.

[11] European Communities, The European Qualifications Framework of Lifelong Learning (EQF). Luxembourg, 2008.

[12] J. P. Trevelyan and S. Tilli, Published Research on Engineering Work. Journal of Professional Issues in Engineering Education and Practice, 133 (4), 300-307, 2007. http://dx.doi.org/10.1061/(ASCE)1052-3928(2007)133:4(300)

[13] VDE. Ingenieurkompetenzen von Berufseinsteigern: Stellungnahme des VDE Ausschusses „Beruf, Gesellschaft und Technik", 2005.

[14] BMBF. Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien zur Förderung von Forschungsvorhaben zum Themenfeld "Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor", 2011.

[15] S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn, and J. Fege (Eds.): Modeling and measuring competencies in higher education. Tasks and challenges. Rotterdam: Sense Publ, 2013. http://dx.doi.org/10.1007/978-94-6091-867-4

[16] H. B. Coates and A. Radloff. Tertiary engineering capability assessment: concept design. Camberwell, Vic.: Australian Council for Educational Research, 2008.

[17] OECD, Assessment of Higher Education Learning Outcomes (Ahelo), 2009.

[18] L. W. Anderson and D. R. Krathwohl, A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives (Abridged). New York u.a.: Longman, 2001.

[19] M. Senkbeil, J. Rost, C. H. Carstensen, and O. Walter. Der nationale Naturwissenschaftstest PISA 2003: Entwicklung und empirische Überprüfung eines zweidimensionalen Facettendesigns. Empirische Pädagogik - Zeitschrift zu Theorie und Praxis erziehungswissenschaftlicher Forschung, 19 (2), pp. 166-189, 2005.

[20] T. Viering, E. H. Fischer, and K. Neumann. Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I: Projekt Physikalische Kompetenz. In E. Klieme, D. Leutner & M. Kenk (Hrsg.) Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms. Zeitschrift für Pädagogik (56). Weinheim: Beltz, pp. 92-102, 2010.

[21] J. S. Bruner, The Process of Education. Cambridge/MASS: Harvard University Press, 1960.

[22] OECD, *Measuring Student Knowledge and Skills: A New Framework for Assessment, 1999.* Retrieved from http://www.oecd.org/dataoecd/45/32/33693997.pdf.

[23] W. Blum, Mathematical modeling in mathematics education and instruction. In T. Breiteig, I. Huntley & G. Kaiser-Messmer (Hrsg.), Teaching and learning mathematics in context. New York: Ellis Horwood, 1993.

[24] M. Neubrand, R. Biehler, W. Blum, E. Cohors-Fresenborg, L. Flade, N. Knoche, et al., Grundlagen der Ergänzung des internationalen PISA Mathematik-Tests in der deutschen Zusatzerhebung: Framework zur Einordnung des PISA Mathematik-Tests in Deutschland: Berichte aus den Expertengruppen - Arbeitsgruppe Mathematik, 1999.

[25] J. Richardson, P. S. Steif, J. Morgan, J., and J. Dantzler, Development of a concept inventory for strength of materials, in *33rd Annual Frontiers in Education, 2003. FIE 2003* (pp. T3D_29). IEEE. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=126334 2.

[26] P. S. Steif, Initial Data from a Statics Concepts Inventory: Proceedings of the 2004, American Society of Engineering Education Conference and Exposition, St. Lake City, UT, June, 2004.

[27] P. S. Steif and J. A. Dantzler, A Statics Concept Inventory: Development and Psychometric Analysis. *Journal of Engineering Education*, *94*(4), 363–371, 2005. http://dx.doi.org/10.1002/j.2168-9830.2005.tb00864.x

[28] P. S. Steif and M. Hansen, Comparisons Between Performances in a Statics Concept Inventory and Course Examinations. *International Journal of Engineering Education*, *22* (1070-1076), 2006.

[29] P. S. Steif and M. Hansen, Feeding Back Results From A Statics Concept Inventory To Improve Instruction: Proceedings of the 2006 American Society of Engineering Education Conference and Exposition, Chicago, IL, June 2006.

[30] P. S. Steif and A. Dollár, Reinventing The Teaching Of Statics. *International Journal of Engineering Education*, *21*, 723–729, 2005.

[31] D. Hestenes, M. Wells, M, and G. Swackhamer, Force Concept Inventory. *The Physics Teacher*, *30* (141-158), 1992. http://dx.doi.org/10.1119/1.2343497

[32] D. Huffman and P. Heller, What does the force concept inventory actually measure? *The Physics Teacher*, *33*, 138–143, 1995. http://dx.doi.org/10.1119/1.2344171

[33] H. Schecker and J. Gerdes, Messung von Konzeptualisierungsfähigkeit in der Mechanik: Zur Aussagekraft des Force Concept Inventory. *Zeitschrift für Didaktik der Naturwissenschaften*, *5*(1), 75–89, 1999.

[34] E. Klieme, M. Neubrand, and O. Lüdke. Mathematische Grundbildung: Testkonstruktion und Ergebnisse. In J. Baumert & M. Neubrand (Hrsg.), PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich, 2001, pp. 139-190.

[35] D. Gross, W. Hauger, J. Schröder, W. A. Wall, and J. Bonet, *Engineering Mechanics 2: Maechanics of Materials*. Berlin: Springer, 2011. http://dx.doi.org/10.1007/978-3-642-12886-8

[36] Hibbeler, R. C. (2010). *Engineering mechanics - statics* (12. ed. in SI units.). Singapore: Prentice Hall, 2010.

[37] G. A. Straka and G. Macke. Neue Einsichten in Lehren- Lernen und Kompetenz. ITB-Forschungsberichte Nr. 40. Bremen: Institut Technik und Bildung, 2009.

[38] K. J. Klauer. Methodik der Lehrzieldefinition und Lehrstoffanalyse. Düsseldorf: Pädagogischer Verlag Schwann, 1974.

[39] Hartig, J.; Frey, A.; Nold, G.; Klieme, E., An Application of Explanatory Item Response Modeling for Model-Based Proficiency Scaling. In: *Educational and Psychological Measurement* 72 (4), S. 665–686, 2012. http://dx.doi.org/10.1177/0013164 411430707

[40] A. Frey, J. Hartig, and A. A. Rupp, Booklet Designs in Large-Scale Assessments of Student Achievement: Theory and Practice. *Educational measurement: issues and practice*, *28*(3), 39–53, 2009. http://dx.doi.org/10.1111/j.1745-3992.2009.00154.x

[41] G. Rasch, Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: MESA Press, 1960.

[42] G. N. Masters, A rasch model for partial credit scoring. Psychometrika, 47(2), 149–174, 1982. http://dx.doi.org/10.1007/BF02296272

[43] M. Kendall and A. Stuart, The advanced theory of statistics. London: Griffin, 1973.

[44] M. Watanabe and K. Yamaguchi. The Em Algorithm and Related Statistical Models: CRC Press, 2004.

[45] M. Garner and G. Engelhard Jr., Using paired comparison matrices to estimate parameters of the partial credit Rasch measurement model for rater-mediated assessments. Journal of Applied Measurement, 10(1), 30–41, 2009.

[46] B. Choppin, A fully conditional estimation procedure for Rasch model parameters. Evaluation in Education, 9(1), 29–42, 1985.

[47] J.-H Heine, and C. Tarnai, Item-Parameter Bestimmung im Rasch-Modell bei unterschiedlichen Datenausfallmechanismen. Referat im 17. Workshop 'Angewandte Klassifikationsanalyse' im Landhaus Rothenberge, oral presentation, 2011.

[48] P. E. Shrout and J. L. Fleiss, Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin*, *86*(2), 420-428, 1979. http://dx.doi.org/10.1037/0033-2909.86.2.420

[49] J.-H. Heine. pairwise: Rasch Model Parameters by Pairwise Algorithm. R package version 0.1.4. retrieved from http://CRAN.R-project.org/package=pairwise, 2013.

[50] M. Bayrhuber, T. Leuders, R. Bruder, and M. Wirtz, Repräsentationswechsel beim Umgang mit Funktionen – Identifikation von Kompetenzprofilen auf der Basis eines Kompetenzstrukturmodells: Projekt Heureko. In E. Klieme, D. Leutner, & M. Kenk (Eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms.* Weinheim: Beltz, 2010.

[51] M. L. Wu, and R. J. Adams, Modelling Mathematics Problemsolving: Item Responses Using a Multidimensional IRT Model. *Mathematics Education Research Journal*, *18*(2), 93–113, 2006. http://dx.doi.org/10.1007/BF03217438

[52] D. R. Sadler, Formative assessment and the design of instructional systems. *Instructional Science*, *18*, pp. 119–144, 1989. http://dx.doi.org/10.1007/BF00117714

[53] Musekamp, F., Mehrafza, M., Heine, J.-H., Schreiber, B., Saniter, A., Spöttl, G., et al. (in press), Formatives Assessment fachlicher Kompetenzen von angehenden Ingenieuren: Validierung eines Kompetenzmodells für die Technische Mechanik im Inhaltsbereich Statik. *Lehrerbildung auf dem Prüfstand (LpP)*, in press.

[54] R. M. Gagné and N. E. Paradise, Abilities and learning sets in knowledge acquisition. *Psychological Monographs: General and Applied*, *75*(14), 1–23, 1961. http://dx.doi.org/10.1037/h0093826

AUTHORS

F. Musekamp is with the Institute Technology and Education (ITB), Bremen University, 28359 Bremen, Germany (e-mail: musekamp@uni-bremen.de).

G. Spöttl is with the Institute Technology and Education (ITB), Bremen University, 28359 Bremen, Germany (e-mail: spoettl@uni-bremen.de).

M. Mehrafza is with Institute for Structural Mechanics and Production Facilities (bime), Bremen University, 28359 Bremen, Germany (e-mail: mehrafza@uni-bremen.de).

J.-H. Heine is with Ludwig-Maximilians-University Munich, Learning Sciences Research Methods, 80802 Munich (e-mail: jhheine@googlemail.com).

M. Heene is with Ludwig-Maximilians-University Munich, Learning Sciences Research Methods, 80802 Munich, (e-mail: heene@psy.lmu.de).