

# Development of a diagnostic instrument for the assessment of teaching competencies in medicine (FKM\_L): First results of the test statistical verification

## Abstract

**Introduction:** This project report describes the development of the *Questionnaire for the Assessment of Teaching Competencies in Medicine* (FKM\_L) and the analysis of some of its psychometric properties. The design of the FKM\_L is based on the model of *Core Competency for Teachers in Medicine* (KLM) model of the *GMA Committee on Personnel and Organizational Development in Teaching*.

**Methods:** Global questions and in-depth items were formulated for each of the sub-competencies of the six core competencies of the KLM model. Depending on the number of sub-competencies, there are 3-4 subscales for each core competency, comprising 69 items in total. Data from 90 participants of medical didactic courses were analysed. Item analyses supported the hypothesized scales.

**Results:** The internal consistencies (Cronbach's alpha: CR- $\alpha$ ) of the 22 subscales ranged from CR- $\alpha$ =.70 to CR- $\alpha$ =.93, and the item difficulty indices of the subscales ranged from 18% to 89%. For 2 subscales, some items had a difficulty index of more than 80%, and for 3 subscales, the difficulty index of some items was less than 25%.

**Conclusions:** The FKM\_L was developed to assess individual and group profiles of teachers' competence in medicine. The results of this first psychometric analysis are promising: With the help of the FKM\_L, teachers can learn about and reflect on aspects of their teaching competencies in the context of medical didactic courses. Based on their FKM\_L profiles, they can decide whether they want to selectively optimize their competence characteristics. For providers, the FKM\_L is suitable as a screening tool to identify, among other things, gaps in the training offered. Further analyses are necessary to check limitations identified in some scales and to improve individual items. In addition, research on the construct and criterion-related validity of the instrument is required.

**Keywords:** teaching competencies in medicine, self-assessment, self-reflection, diagnostic of competencies

Marianne Giesler<sup>1</sup>  
Maria  
Lammerding-Köppel<sup>2</sup>  
Jan Griewatz<sup>3</sup>

1 Freiburg, Germany

2 Tübingen, Germany

3 Kompetenzzentrum für  
Hochschuldidaktik in Medizin  
Baden-Württemberg,  
Tübingen, Germany

## Introduction

### General background

Studies show that a teacher's subject-matter expertise alone is not a sufficient condition for the learning success of individuals being taught [1]. Other subject-independent characteristics and competencies can significantly influence learning development [2]. For school and university settings, conceptual approaches are available that describe competencies teachers should possess in order to cope well with the many and multifaceted tasks and challenges of teaching [3]. For the field of medicine, the *GMA Committee on Personnel and Organizational Development in Teaching* derived a model based on the *Six Core Teaching Competencies for Medical Educators* by

Srinivasan et al. [4] that describes *core competencies for teachers in medicine* (KLM) [2]. This model provides a framework of orientation for medical didactic competencies and indicates what a teacher's qualification profile might look like. For providers of medical didactic qualification programs, the KLM model can be very helpful in orienting their continuing education and training programs in terms of content. However, the KLM model itself cannot yet be used to determine the extent to which individual teachers possess the competencies described and/or to which extent the medical didactic course offerings address these competencies. The core competencies are not described in a sufficiently differentiated manner for this purpose. However, they provide a good basis for developing an appropriate diagnostic tool.

## Diagnosis of the degree of competence

For the diagnosis of the degree of competence or for determining the individual teacher's ability to act in complex (authentic) situations, external and self-assessments can be considered, which can be carried out formatively as well as summatively by means of observations, e.g., by checklists, or by using psychometric procedures [5], [6], [7], [8], [9], [10], [11]. Both methods are subject to random and systematic errors, i.e., the measurement results cannot be perfectly reliable and valid [8]. Self-assessments in particular have a reputation for limited reliability and, accordingly, limited validity [12], [13]: In most cases, evidence is found that those who prove to be least competent are also those who are less accurate in assessing themselves [14], [15]. There is also evidence that highly competent individuals may tend to underestimate their performance [16]. Studies also show that women tend to underestimate their performance [17]. In addition, women's and men's self-assessments may differ depending on the domain of competence [18]. Finally, there is also evidence that practical skills are better assessed than knowledge-based activities [14]. However, self-assessments can be improved by feedback [14], [19]. This can be achieved by providing feedback on the standards to be achieved and offering opportunities for comparison purposes (individual and social) [12], [14], [20]. If the goal is to determine what is required to provide individually suitable offers for further qualification, the aggregation of self-assessment data may allow reliable diagnoses [13].

## Objective

Up to now, there has been no suitable instrument available for the assessment of teaching competencies in medical university teaching that would identify individual and general approaches for the further development of competencies. Therefore, based on the KLM model described above, a diagnostic instrument for the *Assessment of Teaching Competencies in Medicine* (FKM\_L) was developed. This project report describes the development of this instrument and presents first results on its psychometric properties.

## Methods

### Development of the questionnaire

The *Core Competency Model for Teachers in Medicine* (KLM) served as the basis for our development of the *Questionnaire for the Assessment of Teaching Competencies in Medicine* (FKM\_L). In this model, sub-competencies are derived from six fields of competence, for each of which learning objectives and examples of application are also described. The FKM\_L is structured similarly. In a first step, so-called global items were formulated for the individual sub-competencies of the six competence

fields. In a second step, in-depth items were developed for these global items, taking into account the respective learning objectives and examples of application as well as the expertise and experience of the questionnaire developers (see table 1 and table 2), which were then combined into subscales. In formulating the items, attention was paid to conclusiveness and comprehensibility [21]. In addition, participants were asked for feedback on the usefulness of their use of the FKM\_L immediately after having responded to it. This information was included in the further development of the instrument.

The FKM\_L includes 22 global items and 69 in-depth items, each of which is to be responded to on 5-point Likert-type scales ("1" not given at all to "5" given to a very high degree). Only the end points of the scales are semantically anchored.

For the development of the first version of this questionnaire, several meetings of the authors were held in the period from 2018 to 2019. Further revisions were made by way of circulation. Since the middle of 2019, the FKM\_L has been tested in the context of medical didactic offers of the *Competence Center for University Teaching in Medicine Baden-Württemberg in Tübingen*.

## Sample

The analyses were based on the data of two groups of participants of medical didactic courses of the *Competence Center for University Teaching in Medicine Baden-Württemberg*. These competence-oriented courses [22] were held at different faculties in Baden-Württemberg (Freiburg, Mannheim, Tübingen) and included participants mainly from faculties in Baden-Württemberg. In the winter semester 2019/2020, 29 participants of the courses *Medical Teaching Qualification I* (MQ I, basic qualification) and 12 of the courses *Medical Teaching Qualification II* (MQ II, advanced qualification) completed the FKM\_L voluntarily. In the winter semester 2020/2021 and summer semester, there were 49 participants who completed the FKM\_L as part of taking the *Medical Teaching Qualification I* courses.

Because of the small course sizes, we initially refrained from systematically collecting information on age and gender for data protection reasons. In subsequent surveys, this information was obtained on a voluntary basis along with consent forms. Therefore, more detailed information on gender and age is available from only 50 participants. At the time of the survey, these participants were on average 37 years old (SD=5.95, Md=35, Mo=33), 41% are women.

The FKM\_L is used immediately at the beginning of the courses (especially MQ I) as a didactic tool with the aim of introducing the topic of competencies and reflecting on the expressions of one's own teaching competencies. At later points in the course or in the subsequent advanced level, further assessments are carried out. To be able to record individual developments over time, reproducible anonymous codes can be used for the respondents.

**Table 1: Examples of competence domain, global item and in-depth item**

<b>Domain of Competence:</b> Didactical Activities in Medicine	<b>Global Item:</b> You design your lessons in a (learning) goal-oriented way	<b>In-depth Item:</b> You use different teaching/learning methods in your classes
---	--	--

**Table 2: Design and structure of the FKM\_L**

Domains of Competence		Global Items	No. of Items (in-depth)
Didactical Activities in Medicine	MH01	Conception of learning goal oriented lessons	4
	MH02	Design of teaching situation(s) conducive to learning	3
	MH03	Verification of competence development	4
	MH04	Consideration of constructive alignment	2*/3
Student Centred Learning	LO01	Design/use of an atmosphere conducive to learning	3
	LO02	Consideration of prior knowledge	3
	LO03	Support of individual learning process	3
	LO04	Promotion of self-directed learning	3
Social and Communicative Competence	KK01	Comprehensible, structured communication	3
	KK02	Constructive handling of dynamic group processes	3
	KK03	Specification of unambiguous (learning) objectives	3
	KK04	Constructive feedback	3
Role Model and Professional Behaviour	PH01	Reflection on professional actions	3
	PH02	Perception of the function as a role model	3
	PH03	Stimulation to engage with professional action	3
System-based Teaching and Learning	SL01	Course in accordance with curricular context	3
	SL02	Support of curricular processes in faculty	3
	SL03	Active exchange with people outside the faculty	3
	SL04	Consideration of current developments relevant to medicine in curriculum planning	3
Reflection and Further Development of Own Teaching Practice	RW01	Critical review and documentation of teaching behaviour and development.	4
	RW02	Targeted development of teaching competencies	3
	RW03	Continuous expansion of the role spectrum	3

\* Originally, only two items were assigned to this scale

## Statistical analyses

Within the framework of item analyses according to the concept of so-called *Classical Test Theory*, corrected item total correlations as measures of discriminatory power and item difficulty each were calculated for every item. In addition, Cronbach's  $\alpha$  (CR- $\alpha$ ) was calculated for each competence scale as a measure of internal consistency. The concept of item difficulty has originally been introduced in the context of achievement test construction. In this context, item difficulty is estimated using the so-called *difficulty index*. This is defined as the percentage of test takers who solve a test item [23]. It should be noted that a numerically high value of the difficulty index indicates that the task or item in question is "easy" because a high percentage of respondents were able to solve it. Correspondingly, a numerically low difficulty index indicates the presence of a "difficult" item, insofar as it could only be solved by a low percentage of those tested. In order to determine the difficulty of items from self-assessment questionnaires with multiple graded response options, as in the present case, the following formula can

be used to calculate the difficulty index, following Döring and Bortz [24]:  $(M_i - 1 / k - 1) * 100$ , where  $M_i$  denotes the mean of the item in question and  $k$  the number of - coded continuously starting with 1 - levels of the response scale.

As measures of discriminatory power, item total correlations indicate how highly an item correlates with the total score of the associated scale or how well an item measures the construct to be captured by the scale [24], respectively. Corrected item total correlations  $<.30$  are considered low [25]. Values between  $.30$  and  $.50$  are considered medium, and values greater than  $.50$  are considered high.

There is a paraboloid relationship between discriminative power and difficulty [23], i.e., if the difficulty of an item is low, the discriminative power is also low, at an intermediate difficulty of 50% the discriminative power reaches its maximum, thereafter, as the difficulty increases, the discriminative power decreases again. According to Lienert [23], however, medium difficulty does not necessarily mean good discriminatory power.

The following limits for CR- $\alpha$  are generally used to describe reliability levels: high reliability  $>.90$  and low reliability  $<.80$ . However, the meaning of these values may depend on the context and these limits should therefore not be applied too rigidly [24], [25].

Statistical analyses were performed using the statistical program SPSS, version 26.

## Results

### Item analyses

The scale scores determined for the individuals represent sum values that were divided by the number of items in a scale. Accordingly, both the scale scores of the individuals and the means of the scales calculated for the groups can vary between 1 and 5. A high scale score thus means a high level of self-assessed competence in relation to the respective competence field (see table 3).

As shown in table 3, all 22 subscales formed from the in-depth items show internal consistencies above CR- $\alpha=.70$ . Since it is generally recommended to form scales with at least three items, one item was added to the subscale MH04 of the competence domain *Didactical Activities in Medicine* from WS 2020/2021. This led to an increase in the discriminatory power of the items and the reliability of this subscale.

Corrected part-whole correlations between items and their respective scale and difficulty indices were calculated for the items of all subscales. In addition, the mean of the difficulty indices was determined for each subscale. All corrected part-whole correlations are  $>.30$ . Concerning the difficulty indices, the results show that for two competency domains one subscale each (L001, PH02) has a mean difficulty index of  $>80\%$ , i.e. the respective subscales contain rather easy items. In two other competence areas there are also subscales including items with difficulty indices of up to 25% (SL02, RW03), i.e. they are to be regarded as rather difficult.

Further results, which are not presented in detail here, show that the global items, with two exceptions, always correlate most highly and significantly with the subscales assigned to these items. However, similarly high significant correlations are also found with subscales of other competency domains. The correlations of the global items with the associated subscales shown in table 3 range from  $r=.32$  to  $r=.80$ .

A comparison of the mean scores for the global and subscales presented in table 3 shows that the competence domains differ with regard to the level of competence respectively.

### Intercorrelations of the overall scales

The overall scales of the FKM\_L correlate moderately to highly with each other (see table 4). The highest correlations exist between the overall scale *Social and Communicative Competence* (KK) and the overall scale *Student*

*Centred Learning* (LO) as well as between the overall scale *Reflection and Further Development of Own Teaching Practice* (RW) and the overall *System-based Teaching and Learning* (SL) scale.

### Examples of evaluations

To illustrate, figure 1 shows scale scores of the total scales of the six competency domains for each of three randomly selected participants of the MQ I and MQ II courses. Differences between the scale scores can be seen both within and between the competence fields. With one exception, the scale scores of the participants in the MQ II course are all above the scale scores of the participants in the MQ I courses. Accordingly, an analysis of the group means reveals significant differences between the courses for four of the six overall scales.

The illustration of the subscale scores for the three randomly selected participants of the MQ I courses for the competence domain *Social and Communicative Competence* (see figure 2) shows intraindividual and interindividual competence level differences within this competence domain. In the context of counselling events, it can be useful to look at differences of an individual within a subskill at the single-item level. This can be demonstrated by the example of subscale KK04 on constructive feedback for person MQ I\_3, who ticked the three items of this scale as follows:

“1” You follow common feedback rules when giving feedback.

“3” You recognize and use appropriate times for feedback discussions.

“1” You know and use various feedback tools.

These statements may indicate that the person already holds feedback discussions in his or her (teaching) practice, but so far has no knowledge of feedback rules and feedback instruments. In a post-course survey, one would expect these items to be rated more strongly as true. However, a reliable interpretation of such differences requires the determination of appropriate confidence intervals.

## Discussion

The analyses of psychometric properties of the FKM\_L yielded mostly satisfactory results. In the following, the results are discussed separately for the subscales, overall scales, and global items as well as for the examples of evaluations.

### Subscales

For each of the six competence areas, three to four subscales were formed to capture sub-competencies. Since internal consistencies between CR- $\alpha=.60$  and CR- $\alpha=.70$  are considered sufficient for group comparisons [23], [26] and reliabilities of CR- $\alpha=.80$  are considered good

**Table 3: Results of the item analyses for the subscales formed with the in-depth items and for the overall scales for each domain of competence, and correlations between each subscale and the global items of a respective competence domain (N=90)**

Domains of Competence	Subscales (in-depth items)	Scale Mean (SD)	Difficulty Indices % ( $\emptyset$ ) <sup>1</sup>	Item Total Corr. <sup>2</sup> $r_{it}$	CR- $\alpha$	Correlation <sup>3,4</sup> with Global item
Didactical Activities in Medicine	MH01 (4)	3.79 (.65)	65 - 75 (70)	.54 - .65	.78	.55 <sup>a</sup>
	MH02 (3)	3.61 (.83)	53 - 82 (65)	.58 - .63	.77	.39 <sup>c</sup>
	MH03 (4)	2.45 (.89)	30 - 43 (36)	.49 - .71	.82	.41 <sup>b</sup>
	MH04 (2)	3.39 (.83)	59 - 64 (64)	.53	.69	.58 <sup>a</sup>
	MH04 (3) <sup>5</sup>	3.48 (.82)	60 - 64 (62)	.67 - .75	.83	
Overall Scales	MH01 - MH04 <sup>6</sup> (13)	3.27 (.60)		.49 - .63	.86	
Student Centred Learning	LO01 (3)	4.25 (.59)	78 - 88 (81)	.64 - .68	.81	.69 <sup>c</sup>
	LO02 (3)	3.39 (.94)	45 - 70 (60)	.59 - .77	.80	.45 <sup>b</sup>
	LO03 (3)	3.30 (.79)	55 - 59 (57)	.43 - .57	.70	.74 <sup>c</sup>
	LO04 (3)	3.77 (.82)	68 - 71 (69)	.58 - .70	.80	.32 <sup>a</sup>
Overall Scales	LO01 - LO04 (12)	3.68 (.66)		.34 - .65	.88	
Social and Communicative Competence	KK01 (3)	4.05 (.74)	74 - 79 (76)	.66 - .76	.84	.70 <sup>c</sup>
	KK02 (3)	3.28 (.70)	54 - 61 (57)	.63 - .68	.74	.62 <sup>c</sup>
	KK03 (3)	3.83 (.74)	69 - 74 (71)	.70 - .76	.85	.42 <sup>b</sup>
	KK04 (3)	3.46 (.92)	51 - 71 (62)	.62 - .77	.83	.70 <sup>c</sup>
Overall Scales	KK01 - KK04 (12)	3.66 (.63)		.45 - .68	.90	
Role Model and Professional Behaviour	PH01 (3)	3.70 (.86)	65 - 71 (68)	.63 - .76	.83	.80 <sup>c</sup>
	PH02 (3)	4.28 (.69)	77 - 89 (82)	.63 - .67	.81	.69 <sup>c</sup>
	PH03 (3)	3.87 (.70)	66 - 76 (72)	.61 - .76	.83	.64 <sup>c</sup>
Overall Scales	PH01 - PH03 (9)	3.94 (.58)		.45 - .69	.88	
System-based Teaching and Learning	SL01 (3)	3.05 (.88)	40 - 62 (53)	.48 - .65	.74	.55 <sup>c</sup>
	SL02 (3)	2.33 (.97)	18 - 48 (32)	.47 - .71	.75	.75 <sup>c</sup>
	SL03 (3)	2.63 (.90)	22 - 47 (38)	.34 - .67	.71	.75 <sup>c</sup>
	SL04 (3)	3.44 (.94)	51 - 66 (59)	.64 - .72	.81	.60 <sup>a</sup>
Overall Scales	SL01 - SL04 (12)	2.82 (.74)		.28 - .60	.87	
System-based Teaching and Learning	RW01 (4)	3.43 (.86)	47 - 71 (61)	.60 - .70	.83	.76 <sup>c</sup>
	RW02 (3)	2.72 (1.01)	43 (43)	.81 - .91	.93	.73 <sup>c</sup>
	RW03 (3)	2.81 (1.02)	25 - 62 (45)	.46 - .64	.74	.69 <sup>c</sup>
Overall Scales	RW01 - RW03 (10)	3.03 (.79)		.32 - .85	.88	

<sup>1</sup> This column shows the lowest and highest difficulty indices (%) of the individual subscales. The value in parentheses indicates the mean value of the difficulty indices of the items of the respective subscale.

<sup>2</sup> Corrected part-whole correlations

<sup>3</sup> Pearson-correlations, a = <.05, b = <.01, c = p <.000

<sup>4</sup> Global Items were no longer recorded as of WS 2020/2021.

<sup>5</sup> One item was added to the scale for use in WS 20/21 and SoSe 2021

<sup>6</sup> The analyses were conducted with 13 items, as data were available for these from all respondents.

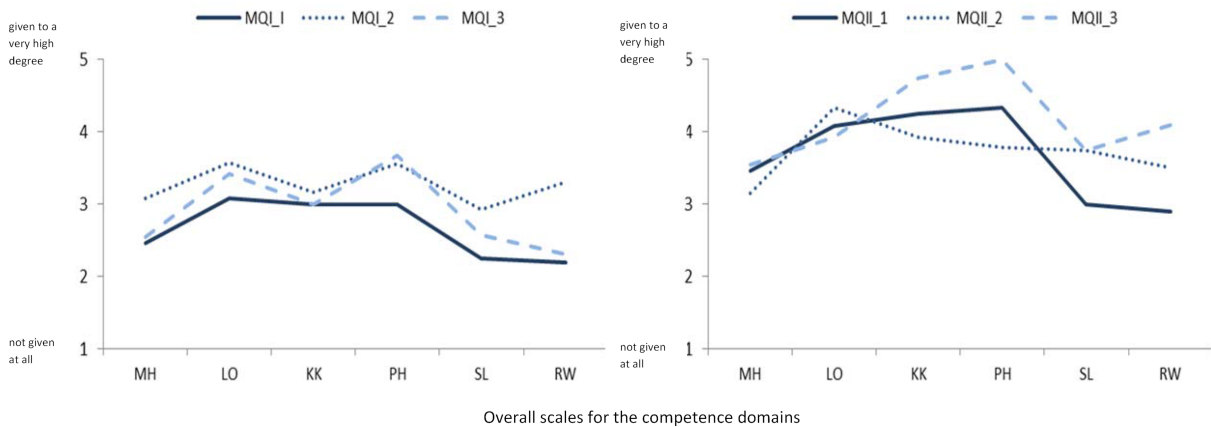
[27], the coefficients (CR- $\alpha$ =.71 to CR- $\alpha$ =.93) obtained for the 22 subscales are in an acceptable to good range. The discriminatory power of an item shows how well the response to the item predicts the corresponding scale sum value [24]. All the discriminatory power values determined for the subscales, with values between .34 and .91, are located in a consistently acceptable to satisfactory range. To determine whether a differentiation between persons with different levels of competence is possible

with the individual items, difficulty indices were calculated for the items of the subscales. In addition, for each subscale the mean of the difficulty indices of their respective items was determined. According to Döring and Bortz [24], extremely difficult items (difficulty index <20%) as well as extremely easy items (difficulty index >80%) are to be regarded as providing little information on their ability to differentiate between respondents. The results show that the latter applies to one subscale each of the

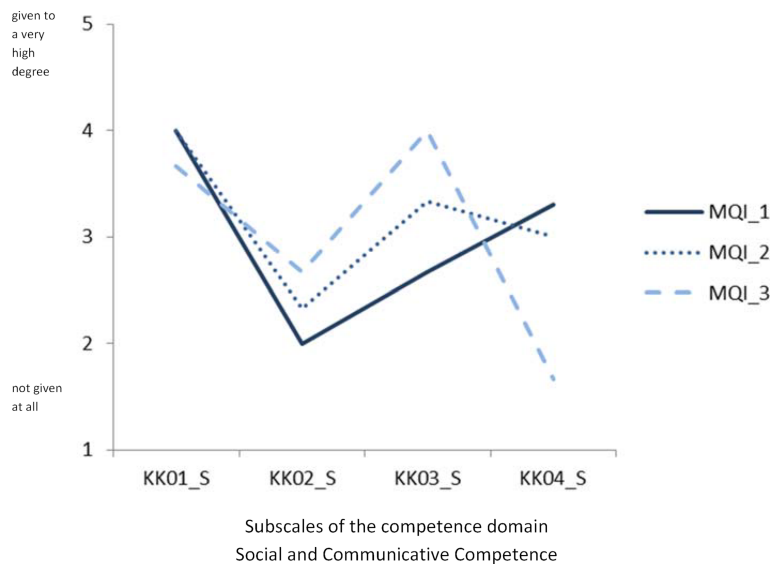
**Table 4: Intercorrelation of the Overall Scales of the FKM\_L**

	Overall Scale_MH	Overall Scale_LO	Overall Scale_KK	Overall Scale_PH	Overall Scale_SL	Overall Scale_RW
Overall Scale_LO	.564 <sup>a</sup>					
Overall Scale_KK	.600 <sup>a</sup>	.713 <sup>a</sup>				
Overall Scale_PH	.634 <sup>a</sup>	.492 <sup>a</sup>	.670 <sup>a</sup>			
Overall Scale_SL	.593 <sup>a</sup>	.317 <sup>a</sup>	.470 <sup>a</sup>	.453 <sup>a</sup>		
Overall Scale_RW	.693 <sup>a</sup>	.350 <sup>a</sup>	.518 <sup>a</sup>	.570 <sup>a</sup>	.799 <sup>a</sup>	

<sup>a</sup> Correlation is significant at the 0.01 level (2-sided); 89 ≤ N ≤ 90



**Figure 1: Scale values for the overall scales of the six competence domains for three different participants in each of the courses MQ I and MQ II**



**Figure 2: Scale values for the subscales of the competence domain Social and Communicative Competence for three participants of the MQ I courses**

competence fields *Student Centred Learning and Role Model and Professional Behaviour*. A closer look at the items suggests that this is due to the fact that the items are phrased in a comparatively global way (e.g. “They value mutual respect”) and may possibly encourage self-serving answers. In three of the subscales, items have difficulty indices between 18% and 25%, i.e. only a maximum of 25 percent of the participants have ticked the items as applicable. Two of these subscales are assigned to the competence field of *System-based Teaching and Learning*, one subscale to the competence field of

*Reflection and Further Development of own Teaching Practice*. The relevant items of these subscales usually address active participation in the planning and implementation of the curriculum. However, the participants of the medical didactics courses are heterogeneous in many ways. The participants come from different subjects, are active in different positions and functions and have been able to gain experience in teaching for more or less time. It is thus understandable that they can assess the items described above as applicable to varying degrees depending on their background. The subscales with too

easy or too difficult items can nevertheless be used as a didactic tool to reflect on the respective teaching competences. Nevertheless, further analysis should be conducted to possibly supplement and/or reformulate the items of these subscales. Additional cognitive pre-tests might provide valuable clues as to how these items could be reworded.

### Overall scales (core competencies)

For each competence domain of the FKM\_L, the items of the subscales were combined into overall scales. These overall scales show consistently good internal consistencies, which is understandable since an extension of scales with additional items usually leads to higher reliability [23].

The FKM\_L was developed based on the model of *Core Competencies for Teachers in Medicine*. Already during the development of this model, it was found that some sub-competencies and learning objectives could be assigned to different competence fields [2]. Therefore, the moderate to high intercorrelations of the total scales can be explained well (see table 4). As determined from the correlations between the total scales, the proportions of their common variance ( $r^2$ ) ranged between .10 and .64. 13 of 15 of these proportions were  $<.50$ , so that it can be assumed that the overall scales capture different facets of teaching competence. However, the very high intercorrelation between the overall scales *Reflection and Further Development of own Teaching Practice* (RW) and *System-based Teaching and Learning* (SL) suggests considering to either combine or shorten these two scales.

### Global items

Most global items correlate significantly, as expected, although sometimes not substantially high ( $r > .50$ ) with their associated subscales (see table 3). In addition, global items of a respective competence domain often also correlate significantly and at comparable levels with the subscales of the other competence domains. This suggests that the global items have low discriminant validity, so that diagnostics on the level of the global items is currently not recommended. For this reason, but also in order to shorten response time and to avoid negative effects possibly associated with response load, these items were dispensed with from WS 2020/2021 onward.

### Examples of evaluations

With the examples of individual evaluations given, inter- and intra-individual differences in the competency characteristics are apparent. As a next step, comparison values of appropriate reference groups will be provided. However, this requires a broader data base.

### Feedback on the questionnaire

The feedback of the participants of the different courses referred predominantly to the length of the questionnaire, i.e., to the number of items and the length of the wording of individual items. It was suggested on various occasions that the completion time should be increased to allow a more thorough assessment of one's competencies. In addition, the use of terms not understandable without further prior information was criticized. The manner of presentation (especially the assignment of items to competencies vs. arbitrary arrangement; size of screen pages) was also addressed. Short additional information in the form of marginal notes on the reading of certain items was desired in order to make the assessments more reliable.

### Validity

So far, steps have only been taken to ensure the content validity of the FKM\_L by taking the KLM model as the basis for development. Furthermore, the authors wrote and assigned the items by drawing on theoretical arguments [24], [25]. However, a thorough examination of the construct and criterion validity of the instrument is still pending.

### Conclusion

In summary, based on the presented results, the FKM\_L can already be used for individual and group diagnostics despite slight limitations: Teachers can use their FKM\_L profiles to decide whether they want to review or optimize their competency characteristics in one respect or another, and faculty developers can use the FKM\_L as a screening tool to identify participants' needs and gaps in their qualification offerings. In addition, the assessment of competencies and their results can be used as a didactic tool in medical didactic courses, e.g., to raise awareness and reflect on individual teaching competencies and their development needs.

### Funding

The study was conducted within the context of the BMBF-funded joint project MERLIN (Medical Education Research – *Lehrforschung im Netz BW*) of the medical faculties of Freiburg, Heidelberg, Mannheim, Ulm and Tübingen under the leadership of the Freiburg and Tübingen sites. Grant number TÜ: O1PL17011A, FR: O1PL17011B

### Competing interests

The authors declare that they have no competing interests.

## References

- Bromme R. Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In: Weinert FE, editor. Enzyklopädie der Psychologie. Psychologie des Unterrichts und der Schule. Göttingen: Hogrefe; 1997. p.177-212.
- Görlitz A, Ebert T, Bauer D, Grasl M, Hofer M, Lammerding-Köppel M, Fabry G. Core Competencies for Medical Teachers (KLM) – A Position Paper of the GMA Committee on Personal and Organizational Development in Teaching. *GMS Z Med Ausbild.* 2015;32(2):Doc23. DOI: 10.3205/zma000965
- Lipowsky F. Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In: Allemann-Ghionda C, Terhart E, editors. Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf. Weinheim: Beltz; 2006. p.47-70.
- Srinivasan M, T Li ST, Meyers FJ, Pratt DD, Collins JB, Braddock C, Hilty DM. "Teaching as a Competency": competencies for medical educators. *Acad Med.* 2011;86(10):1211-1220. DOI: 10.1097/ACM.0b013e31822c5b9a 13
- Braun E, Athanassios G, Pollerhof K, Schwabe U. Wie lassen sich kommunikative Kompetenzen messen? Konzeption einer kompetenzorientierten Prüfung kommunikativer Fähigkeiten von Studierenden. *Beitr Hochschulforsch.* 2018;40(3):34-55. Zugänglich unter/available from: <https://www.fachportal-paedagogik.de/literatur/vollanzeige.html?Fid=1148973>
- Braun E, Gusy B, Leidner B, Hannover B. Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica.* 2008;54(1):30-43. DOI: 10.1026/0012-1924.54.1.30
- Giesler M, Forster J, Biller S, Fabry G. Entwicklung eines Fragebogens zur Erfassung von Kompetenzen in der Medizin: Ergebnisse zur Reliabilität und Validität. *GMS Z Med Ausbild.* 2011;28(2):Doc31. DOI: 10.3205/zma000743
- Gollwitzer M, Jäger RS. Evaluation – Workshop. Weinheim: Beltz PVU; 2007.
- Jurecka A. Introduction of the Computer-Based Assessment of Competencies. In: Hartig J, Klieme E, Leutner D, editors. Assessment of Competencies in Educational Contexts. Bern: Hogrefe; 2008. p.193-213.
- Hartig J. Psychometric models for the assessment of competencies. In: Hartig J, Klieme E, Leutner D, editors. Assessment of Competencies in Educational Contexts. Bern: Hogrefe; 2008. p.69-90.
- Wass V, Vleuten van der C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet.* 2001;357(9260):945-949. DOI: 10.1016/S0140-6736(00)04221-5
- Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med.* 2005;80(10 suppl):S46-54. DOI: 10.1097/00001888-200510001-00015
- Peterson LN, Eva KW, Rusticus SA. The readiness for Clerkship Survey: Can Self-assessment Data be used to evaluate Program Effectiveness? *Acad Med.* 2012;87(10):1355-1360. DOI: 10.1097/ACM.0b013e3182676c76
- Colthart I, Bagnell G, Evans A, Allbutt H, Haig A, Illing J, Mckinstry B. The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. *Med Teach.* 2008;30(2):124-145. DOI: 10.1080/01421590701881699
- Jünger J, Schellberg D, Nikendei C. Subjektive Kompetenzeinschätzung von Studierenden und ihre Leistung im OSCE. *GMSZ Med Ausbild.* 2006;23(3):Doc51. Zugänglich unter/available from: <https://www.egms.de/static/de/journals/zma/2006-23/zma000270.shtml>
- Sieverding M. Frauen unterschätzen sich: Selbstbeurteilungs-Biases in einer simulierten Bewerbungssituation. *Z Sozialpsychol.* 2003;34(3):147-160. DOI: 10.1024//0044-3514.34.3.147
- León SP, Vallejo AP, Nelson JB. Variability in the accuracy of self-assessments among low, moderate, and high performing students in university education. *Pract Ass Res Evaln.* 2021;26:Article 16. DOI: 10.7275/6q91-az58
- Bußenius L, Harendza S, van den Bussche H, Selch S. Final-year medical students' self-assessment of facets of competence for beginning residents. *BMC Med Educ.* 2022;22(1):82. DOI: 10.1186/s12909-021-03039-2
- Ehrlinger J, Dunning D. How chronic selfviews influence (and potentially mislead) estimates of performance. *J Pers Soc Psychol.* 2003;84(1): 5-17. DOI: 10.1037/0022-3514.84.1.5
- Regehr G, Hodges B, Tiberius R, Lofchy J. Measuring self-assessment skills: An innovative relative ranking model. *Acad Med.* 1996;71(10 Suppl):852-854. DOI: 10.1097/00001888-199610000-00043
- Schulz von Thun F. Miteinander reden 1: Störungen und Klärungen. Allgemeine Psychologie der zwischenmenschlichen Kommunikation. 53. Auflage. Reinbek (bei Hamburg): Rowohlt-TB; 1981.
- Griewatz J, Simon M, Lammerding-Koeppel M. Competency-based teacher training: A systematic revision of a proven programme in medical didactics. *GMS J Med Educ.* 2017;34(4):Doc44. DOI: 10.3205/zma001121
- Lienert GA. Testaufbau und Testanalyse. 2. Auflage. Weinheim: Beltz; 1967.
- Döring N, Bortz J. Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. 5. Auflage. Heideberg: Springer; 2016. DOI: 10.1007/978-3-642-41089-5
- Bühner M. Einführung in die Test- und Fragebogenkonstruktion. 3. akt. Auflage. München: Pearson Studium; 2011.
- Rammstedt B. Zur Bestimmung der Güte von Multi-Item-Skalen: Eine Einführung (ZUMA How-to-Reihe Nr. 12). Mannheim: ZUMA; 2004.
- Nunnally JC, Bernstein IH. Psychometric Theory. 3. Ed. New York: McGraw-Hill; 1994.

### Corresponding author:

Dr. Marianne Giesler  
Freiburg, Germany  
Dr\_M\_Giesler@t-online.de

### Please cite as

Giesler M, Lammerding-Köppel M, Griewatz J. Development of a diagnostic instrument for the assessment of teaching competencies in medicine (FKM\_L): First results of the test statistical verification. *GMS J Med Educ.* 2022;39(5):Doc53. DOI: 10.3205/zma001574, URN: urn:nbn:de:0183-zma001574

This article is freely available from  
<https://doi.org/10.3205/zma001574>

Received: 2022-01-05  
Revised: 2022-07-28  
Accepted: 2022-08-04  
Published: 2022-11-15



**Copyright**

©2022 Giesler et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

# Entwicklung eines Diagnoseinstruments zur Erfassung von Lehrkompetenzen in der Medizin (FKM\_L): Erste Ergebnisse der teststatistischen Überprüfung

## Zusammenfassung

**Einführung:** In diesem Projektbericht wird die Entwicklung und teststatistische Überprüfung des *Fragebogens zur Erfassung von Lehrkompetenzen in der Medizin* (FKM\_L) beschrieben. Die Basis für die Konzeption des FKM\_L bildete das Modell *Kernkompetenzen für Lehrende in der Medizin* (KLM) des *GMA Ausschusses Personal-/Organisationsentwicklung in der Lehre*.

**Methodik:** Für die Teilkompetenzen der sechs Kernkompetenzen des KLM-Modells wurden jeweils Globalfragen und vertiefende Items formuliert. Je Kernkompetenz gibt es in Abhängigkeit von der Zahl der Teilkompetenzen 3-4 Subskalen, die 69 Items umfassen. Alle Skalen wurden itemanalytisch überprüft. Analysiert wurden Daten von 90 Teilnehmerinnen und Teilnehmern medizindidaktischer Kurse.

**Ergebnisse:** Die inneren Konsistenzen (Cronbachs Alpha: CR- $\alpha$ ) der 22 Subskalen liegen zwischen CR- $\alpha$ =.70 und CR- $\alpha$ =.93, die Schwierigkeitsindices der Items der Subskalen zwischen 18% und 89%. Bei 2 Subskalen wiesen einige Items einen Schwierigkeitsindex von über 80% auf, bei 3 Subskalen betrug der Schwierigkeitsindex einiger Items weniger als 25%.

**Schlussfolgerung:** Der FKM\_L wurde zur Erfassung von individuellen und gruppenbezogenen Kompetenzprofilen von Lehrenden in der Medizin entwickelt. Die Ergebnisse dieser ersten teststatistischen Analyse sind vielversprechend: Mit Hilfe des FKM\_L können Lehrende im Rahmen medizindidaktischer Kurse Aspekte ihrer Lehrkompetenzen kennenlernen und reflektieren. Anhand ihrer FKM\_L-Profile können sie entscheiden, ob sie ihre Kompetenzausprägungen punktuell optimieren möchten. Für Anbieter eignet sich der FKM\_L als Screening-Instrument, um u.a. Lücken im Ausbildungsangebot festzustellen. Weitere Analysen sind notwendig, um festgestellte Einschränkungen bei einigen Skalen zu überprüfen und einzelne Items anzupassen. Darüber hinaus ist eine Überprüfung der Konstrukt- und der kriterienbezogenen Validität des Instruments erforderlich.

**Schlüsselwörter:** Lehr-Kompetenzen in der Medizin, Selbsteinschätzungen, Selbstreflexion, Kompetenzdiagnostik

## Einleitung

### Allgemeiner Hintergrund

Studien zeigen, dass die fachliche Expertise einer Lehrperson alleine keine hinreichende Voraussetzung für den Lernerfolg von Personen ist, die unterrichtet werden [1]. Weitere fachunabhängige Merkmale und Kompetenzen können erheblichen Einfluss auf die Lernentwicklung haben [2]. Es existieren für den schulischen und universitären Bereich Ansätze, die Kompetenzen von Lehrkräften beschreiben, über die diese verfügen sollten, um die

vielen und vielschichtigen Aufgaben und Herausforderungen gut bewältigen zu können [3]. Für den Bereich Medizin wurde vom *GMA Ausschuss für Personal- und Organisationsentwicklung in der Lehre* auf der Basis der *Six Core Teaching Competencies for Medical Educators* von Srinivasan et al. [4] ein Modell abgeleitet, das *Kernkompetenzen für Lehrende in der Medizin* (KLM) beschreibt [2]. Dieses Modell stellt einen Orientierungsrahmen für medizindidaktische Kompetenzen zur Verfügung und zeigt auf, wie ein Qualifikationsprofil für Lehrende aussehen könnte. Für Anbieter von medizindidaktischen Qualifikationsangeboten kann das KLM-Modell sehr hilfreich sein, um ihre Weiter- und Fortbildungen inhaltlich auszurichten.

Marianne Giesler<sup>1</sup>

Maria

Lammerding-Köppel<sup>2</sup>

Jan Griewatz<sup>3</sup>

1 Freiburg, Deutschland

2 Tübingen, Deutschland

3 Kompetenzzentrum für  
Hochschuldidaktik in Medizin  
Baden-Württemberg,  
Tübingen, Deutschland

Mit dem KLM-Modell selbst kann jedoch noch keine Standortbestimmung vorgenommen werden, um festzustellen, inwieweit einzelne Lehrkräfte über die beschriebenen Kompetenzen verfügen und/oder die medizindidaktischen Kursangebote diese Kompetenzen adressieren. Dazu sind die Kernkompetenzen nicht differenziert genug beschrieben. Sie bieten jedoch eine gute Grundlage für die Entwicklung eines entsprechenden Diagnoseinstruments.

## Diagnostik von Kompetenzausprägungen

Zur Diagnose von Kompetenzausprägungen oder zur Feststellung des Ausmaßes der individuellen Handlungsfähigkeit in komplexen (authentischen) Situationen kommen Fremd- und Selbsteinschätzungen in Betracht, die mittels Beobachtungen z.B. anhand von Checklisten oder durch den Einsatz von psychometrischen Verfahren sowohl formativ wie summativ erfolgen können [5], [6], [7], [8], [9], [10], [11]. Beide Methoden sind mit unsystematischen und systematischen Fehlern behaftet, d.h. die Messergebnisse können nicht perfekt reliabel und valide sein [8]. Vor allem den Selbsteinschätzungen haftet der Ruf an, eingeschränkt reliabel und dementsprechend auch begrenzt valide zu sein [12], [13]: Meist finden sich Belege dafür, dass die, die sich am wenigsten kompetent erweisen, auch diejenigen sind, die sich weniger genau einschätzen können [14], [15]. Es gibt zudem Belege dafür, dass sehr kompetente Personen dazu neigen können, ihre Leistungen zu unterschätzen [16]. Auch Frauen tendieren dazu, ihre Leistungen zu unterschätzen [17]. Darüber hinaus können sich die Selbsteinschätzungen von Frauen und Männer in Abhängigkeit von der jeweiligen Kompetenzdomäne unterscheiden [18]. Es gibt zudem Hinweise, dass praktische Fertigkeiten besser eingeschätzt werden als wissensbasierte Aktivitäten [14]. Selbsteinschätzungen können jedoch durch Feedback verbessert werden [14], [19]. Dies lässt sich dadurch erreichen, dass Rückmeldungen über die zu erreichenden Standards zur Verfügung gestellt und Vergleichsmöglichkeiten (individuelle und soziale) angeboten werden [12], [14], [20]. Wenn zur Bereitstellung passender Qualifikationsangebote Standortbestimmungen bzw. Bedarfsanalysen benötigt werden, kann die Aggregation von Selbsteinschätzungsdaten durchaus reliable Diagnosen ermöglichen [13].

## Zielsetzung

Für die medizinische Hochschullehre steht bislang kein geeignetes Instrument zur Erfassung von Lehrkompetenzen zur Verfügung, um individuelle und generelle Ansatzpunkte zur Weiterentwicklung der Kompetenzen zu identifizieren. Es wurde daher ausgehend von dem oben beschriebenen KLM-Modell ein Diagnoseinstrument zur *Erfassung von Lehrkompetenzen in der Medizin* (FKM\_L) entwickelt. Dieser Projektbericht beschreibt die Entwicklung dieses Instruments und stellt erste teststatistische Ergebnisse dar.

## Methodik

### Entwicklung des Fragebogens

Als Basis für die Entwicklung des *Fragebogens zur Erfassung von Lehrkompetenzen in der Medizin* (FKM\_L) diente das *Kernkompetenzmodell für Lehrende in der Medizin* (KLM). In diesem Modell werden ausgehend von sechs Kompetenzfeldern jeweils Teilkompetenzen abgeleitet, für die ebenfalls jeweils Lernziele und Anwendungsbeispiele beschrieben werden. Der FKM\_L ist ähnlich aufgebaut. In einem ersten Schritt wurden für die einzelnen Teilkompetenzen der sechs Kompetenzfelder sogenannte Globalitems formuliert. In einem zweiten Schritt wurden für diese Globalitems unter Berücksichtigung der jeweiligen Lernziele und Anwendungsbeispiele sowie der Expertise und den Erfahrungen der Fragebogenentwicklerinnen und -entwickler vertiefende Items entwickelt (siehe Tabelle 1 und Tabelle 2), die anschließend zu Subskalen zusammengefasst wurden. Bei der Formulierung der Items wurde auf inhaltliche Schlüssigkeit und Verständlichkeit [21] geachtet. Zudem wurden die Teilnehmerinnen und Teilnehmer unmittelbar nach Beantwortung des FKM\_L um Rückmeldung zum Nutzen seines Einsatzes gebeten. Diese Angaben wurden in den Weiterentwicklungsprozess des Instrumentes einbezogen.

Der FKM\_L besteht aus 22 Globalitems und 69 vertiefenden Items, die jeweils anhand 5-stufiger Skalen zu beantworten sind („1“ gar nicht gegeben bis „5“ in sehr hohem Maße gegeben). Nur die Endkategorien der Skalen sind semantisch verankert.

Für die Erstellung der ersten Version dieses Fragebogens fanden im Zeitraum von 2018 bis 2019 mehrere Arbeitstreffen der beiden Autorinnen und des Autors statt. Die weiteren Überarbeitungen erfolgten im Umlaufverfahren. Seit Mitte 2019 wird der FKM\_L im Rahmen von medizindidaktischen Angeboten des *Kompetenzzentrums für Hochschuldidaktik in der Medizin Baden-Württemberg in Tübingen* erprobt.

### Stichprobe

Den Analysen lagen die Daten von zwei Gruppen von Teilnehmerinnen und Teilnehmern medizindidaktischer Kurse des *Kompetenzzentrums für Hochschuldidaktik in der Medizin Baden-Württemberg* zugrunde. Diese kompetenzorientiert gestalteten Kurse [22] wurden an verschiedenen baden-württembergischen Fakultäten (Freiburg, Mannheim, Tübingen) unter Einbezug von Teilnehmerinnen und Teilnehmern v.a. aus baden-württembergischen Fakultäten durchgeführt. Im Wintersemester 2019/2020 haben 29 Teilnehmerinnen und Teilnehmer der Kurse *Medizindidaktische Qualifikation I* (MQ I, Basisqualifikation) und 12 der Kurse *Medizindidaktische Qualifikation II* (MQ II, Aufbauqualifikation) den FKM\_L freiwillig ausgefüllt. Im Wintersemester 2020/2021 und Sommersemester waren es 49 Personen, die den FKM\_L im Rahmen ihrer Teilnahme an den *Medizindidaktischen Qualifikationskursen I* bearbeitet haben.

Tabelle 1: Beispiele für Kompetenzfeld, Globalitem und vertiefendes Item

<b>Kompetenzfeld:</b> Medizindidaktisches Handeln	<b>Globalitem:</b> Sie konzipieren Ihren Unterricht (lern-)zielorientiert	<b>Vertiefendes Item:</b> Sie nutzen in Ihrem Unterricht unterschiedliche Lehr-/Lernmethoden
--	--	---

Tabelle 2: Aufbau und Struktur des FKM\_L

Kompetenzfelder		Global-Items	Anz. Items (vertiefend)
Medizindidaktisches Handeln	MH01	Konzeption lernzielorientierter Unterricht	4
	MH02	Gestaltung lernförderlicher Unterrichtssituation/en	3
	MH03	Überprüfung Kompetenzentwicklung	4
	MH04	Berücksichtigung des Constructive Alignment	2*/3
Lernerorientierung	LO01	Gestaltung/Nutzung lernförderlicher Atmosphäre	3
	LO02	Berücksichtigung Vorwissen	3
	LO03	Unterstützung individueller Lernprozess	3
	LO04	Förderung selbstgesteuertes Lernen	3
Soziale und Kommunikative Kompetenz	KK01	Verständliche, strukturierte Kommunikation	3
	KK02	Konstruktiver Umgang mit gruppenspezifischen Prozessen	3
	KK03	Vorgabe von eindeutigen (Lern-)Ziele	3
	KK04	Konstruktives Feedback	3
Rollenvorbild und professionelles Handeln	PH01	Reflexion professionellen Handelns	3
	PH02	Wahrnehmung der Vorbildfunktion	3
	PH03	Anregung zur Auseinandersetzung mit professionellem Handeln	3
Systembezogenes Lehren und Lernen	SL01	Lehrveranstaltung im Einklang mit curricularem Kontext	3
	SL02	Curriculare Prozesse in Fakultät unterstützen	3
	SL03	Aktiver Austausch mit Personen außerhalb der Fakultät	3
	SL04	Berücksichtigung aktueller medizinrelevanter Entwicklungen in Lehrplanung	3
Reflexion und Weiterentwicklung der eigenen Lehrpraxis	RW01	Kritische Überprüfung und Dokumentation des Lehrverhaltens und der Entwicklung	4
	RW02	Gezielte Weiterentwicklung von Lehrkompetenzen	3
	RW03	Kontinuierlicher Ausbau des Rollenspektrums	3

\*Ursprünglich waren dieser Skala nur zwei Items zugeordnet

Wegen der kleinen Kursgrößen haben wir aus Datenschutzgründen zunächst auf eine systematische Erfassung der Angaben zu Alter und Geschlecht verzichtet. Bei späteren Erhebungen wurden diese Informationen auf freiwilliger Basis zusammen mit den Einverständniserklärungen eingeholt. Es liegen daher von nur 50 Personen nähere Angaben zu Geschlecht und Alter vor. Zum Zeitpunkt der Befragung waren diese im Mittel 37 Jahre (SD=5.95, Md=35, Mo=33) alt, 41% waren Frauen. Der FKM\_L wird unmittelbar zu Beginn der Kurse (v.a. MQ I) als didaktisches Mittel mit dem Ziel der Einführung in die Thematik Kompetenzen und zur Reflexion der Ausprägungen der eigenen Lehrkompetenzen eingesetzt. Zu späteren Zeitpunkten im Kurs oder in der anschließenden Aufbaustufe werden weitere Standortbestimmungen durchgeführt. Um individuelle Entwicklungen im Zeitverlauf erfassen zu können, können reproduzierbare anonyme Codes für die Befragten verwendet werden.

## Statistische Analysen

Im Rahmen von Itemanalysen nach dem Konzept der sog. *Klassischen Testtheorie* wurden für jedes Item Trennschärfe und Schwierigkeit berechnet. Als Maß der inneren Konsistenz wurde für jede Kompetenz-Skala zu dem Cronbach's  $\alpha$  (CR- $\alpha$ ) ermittelt.

Das Konzept der Schwierigkeit von Items wurde ursprünglich im Zusammenhang der Konstruktion von Leistungstests eingeführt. Die Itemschwierigkeit wird dabei mit Hilfe des sog. Schwierigkeitsindex geschätzt. Dieser ist definiert als der Prozentsatz der Getesteten, die eine Testaufgabe bzw. ein Testitem löst [23]. Hierbei ist zu beachten, dass ein numerisch hoher Wert des Schwierigkeitsindex anzeigt, dass die betreffende Aufgabe bzw. das betreffende Item „leicht“ ist, da ein hoher Prozentsatz der Befragten sie lösen konnte. Entsprechend zeigt ein numerischer niedriger Schwierigkeitsindex das Vorliegen eines „schwierigen“ Items an, insofern dieses nur von einem geringen Prozentsatz der Getesteten gelöst werden konnte. Um die Schwierigkeit von Items aus Selbsteinschätzungsfragebögen mit mehrfach abgestuften Antwort-

möglichkeiten wie im hier vorliegenden Fall bestimmen zu können, kann in Anlehnung an Döring und Bortz [24] zur Berechnung des Schwierigkeitsindex die folgende Formel herangezogen werden:  $(M_i - 1 / k - 1) * 100$ . Dabei bezeichnet  $M_i$  den Mittelwert des fraglichen Items und  $k$  die Anzahl der – beginnend mit 1 kontinuierlich kodierten – Stufen der Antwortskala.

Trennschärfekoeffizienten geben an, wie hoch ein Item mit dem Gesamtwert der zugehörigen Skala korreliert bzw. wie gut ein Item das von der Skala zu erfassende Konstrukt misst [24]. Trennschärfen  $<.30$  werden als niedrig eingestuft [25]. Als „mittlere“ Trennschärfen gelten Werte zwischen  $.30$  und  $.50$ , Werte größer als  $.50$  als hohe Trennschärfen.

Zwischen Trennschärfe und Schwierigkeit besteht eine paraboloid Beziehung [23], d.h. bei geringer Schwierigkeit eines Items ist auch die Trennschärfe gering, bei einer mittleren Schwierigkeit von 50% erreicht die Trennschärfe ihr Maximum, danach nimmt bei steigender Schwierigkeit die Trennschärfe wieder ab. Nach Lienert [23] bedeutet mittlere Schwierigkeit jedoch nicht zwangsläufig eine gute Trennschärfe.

Zur Beurteilung der Reliabilität werden im Allgemeinen folgende Wertegrenzen für CR- $\alpha$  herangezogen: hohe Reliabilität  $>.90$  und niedrige Reliabilität  $<.80$ . Diese Größen sind jedoch kontextabhängig und sollten daher nicht zu starr angewandt werden [24], [25].

Die statistischen Auswertungen erfolgten mit Hilfe des Statistikprogramms SPSS, Version 26.

## Ergebnisse

### Itemanalysen

Die für die Personen ermittelten Skalenwerte stellen Summenwerte dar, die durch die Anzahl der Items einer Skala dividiert wurden. Entsprechend können sowohl die Skalenwerte der Personen als auch die für die Gruppen berechneten Mittelwerte der Skalen zwischen 1 und 5 variieren. Ein hoher Skalenwert bedeutet somit eine hohe Ausprägung der selbsteingeschätzten Kompetenz in Bezug auf das jeweilige Kompetenzfeld (siehe Tabelle 3).

Wie aus Tabelle 3 hervorgeht, weisen alle 22 aus den vertiefenden Items gebildeten Subskalen Maße der inneren Konsistenz auf, die über CR- $\alpha=.70$  liegen. Da allgemein empfohlen wird, Skalen mit mindestens drei Items zu bilden, wurde ab WS 2020/2021 die Subskala MH04 des Kompetenzfeldes Medizindidaktisches Handeln um ein Item ergänzt. Dies führte zu einer Erhöhung der Trennschärfen und der Reliabilität dieser Subskala.

Für die Items aller Subskalen wurden (part-whole korrigierte) Trennschärfen und Schwierigkeitsindices berechnet. Darüber hinaus wurde pro Subskala jeweils der Mittelwert der Schwierigkeitsindices bestimmt. Die Trennschärfen sind alle  $>.30$ . Bezüglich der Schwierigkeitsindices zeigen die Ergebnisse, dass für zwei Kompetenzfelder jeweils eine Subskala (L001, PH02) einen mittleren Schwierigkeitsindex von  $>80\%$  aufweist, die betreffenden

Subskalen also eher leichte Items beinhalten. Bei zwei weiteren Kompetenzfeldern finden sich zudem Subskalen, deren Items Schwierigkeitsindices von bis zu 25% (SL02, RW03) aufweisen, also eher als schwierig anzusehen sind.

Weitere hier nicht näher dargestellte Ergebnisse zeigen, dass die Globalitems, von zwei Ausnahmen abgesehen, immer am höchsten und signifikant mit den diesen Items zugeordneten Subskalen korrelieren. Es finden sich jedoch auch ähnlich hohe signifikante Zusammenhänge mit Subskalen anderer Kompetenzfelder. Die in Tabelle 3 dargestellten Korrelationen der Globalitems mit den dazugehörigen Subskalen liegen zwischen  $r=.32$  und  $r=.80$ .

Ein Vergleich der in Tabelle 3 dargestellten Mittelwerte für die Gesamt- und Subskalen zeigt, dass die Kompetenzfelder sich im Hinblick auf das jeweils vorhandene Kompetenzniveau unterscheiden.

### Interkorrelationen der Gesamtskalen

Die Gesamtskalen des FKM\_L korrelieren moderat bis hoch miteinander (siehe Tabelle 4). Die höchsten Korrelationen bestehen zwischen der Gesamtskala *Kommunikative Kompetenz* (KK) und der Gesamtskala *Lernorientierung* (LO) sowie zwischen der Gesamtskala *Reflexion und Weiterentwicklung der eigenen Lehrpraxis* (RW) und der Gesamtskala *Systembezogenes Lehren und Lernen* (SL).

### Auswertungsbeispiele

In Abbildung 1 sind jeweils für drei zufällig ausgewählte Teilnehmerinnen und Teilnehmer der MQ I und MQ II-Kurse exemplarisch die Skalenwerte der Gesamtskalen der sechs Kompetenzfelder dargestellt. Unterschiede zwischen den Skalenwerten sind sowohl innerhalb als auch zwischen den Kompetenzfeldern zu erkennen. Die Skalenwerte der Teilnehmerinnen und Teilnehmer des MQ II-Kurses liegen bis auf eine Ausnahme alle über den Skalenwerten der Teilnehmerinnen und Teilnehmer der MQ I-Kurse. Eine Analyse der Gruppenmittelwerte ergibt dementsprechend für vier der sechs Gesamtskalen signifikante Unterschiede zwischen den Kursen.

Die exemplarische Darstellung der Subskalenwerte für die drei zufällig ausgewählten Teilnehmerinnen und Teilnehmer der MQ I-Kurse für das Kompetenzfeld *Soziale und Kommunikative Kompetenz* (siehe Abbildung 2) zeigt intraindividuelle und interindividuelle Kompetenzniveauunterschiede innerhalb dieses Kompetenzfeldes. Im Rahmen von Beratungsanlässen kann es sinnvoll sein, Unterschiede einer Person innerhalb einer Teilkompetenz auf Einzelitemebene zu betrachten. Dies lässt sich am Beispiel der Subskala KK04 zum konstruktiven Feedback für die Person MQ I\_3 veranschaulichen, die die drei Items dieser Skala wie folgt angekreuzt hat:

Tabelle 3: Ergebnisse der Itemanalysen für die mit den vertiefenden Items gebildeten Subskalen und für die Gesamtskalen der Kompetenzfelder sowie Korrelationen der Subskalen mit den Globalitems des jeweiligen Kompetenzfeldes (N=90)

Kompetenzfelder	Subskalen (vertief. Items)	Skalen-mittel-wert (SD)	Schwierigkeits-indices % (Ø) <sup>1</sup>	Trennschärfen <sup>2</sup> r <sub>tt</sub>	CR-α	Korrelation <sup>3,4</sup> mit Globalitem
Medizindidaktisches Handeln	MH01 (4)	3.79 (.65)	65 - 75 (70)	.54 - .65	.78	.55 <sup>a</sup>
	MH02 (3)	3.61 (.83)	53 - 82 (65)	.58 - .63	.77	.39 <sup>c</sup>
	MH03 (4)	2.45 (.89)	30 - 43 (36)	.49 - .71	.82	.41 <sup>b</sup>
	MH04 (2) MH04 (3) <sup>5</sup>	3.39 (.83) 3.48 (.82)	59 - 64 (64) 60 - 64 (62)	.53 .67 - .75	.69 .83	.58 <sup>a</sup>
Gesamtskala	MH01 - MH04 <sup>6</sup> (13)	3.27 (.60)		.49 - .63	.86	
Lernerorientierung	LO01 (3)	4.25 (.59)	78 - 88 (81)	.64 - .68	.81	.69 <sup>c</sup>
	LO02 (3)	3.39 (.94)	45 - 70 (60)	.59 - .77	.80	.45 <sup>b</sup>
	LO03 (3)	3.30 (.79)	55 - 59 (57)	.43 - .57	.70	.74 <sup>c</sup>
	LO04 (3)	3.77 (.82)	68 - 71 (69)	.58 - .70	.80	.32 <sup>a</sup>
Gesamtskala	LO01 - LO04 (12)	3.68 (.66)		.34 - .65	.88	
Soziale und Kommunikative Kompetenz	KK01 (3)	4.05 (.74)	74 - 79 (76)	.66 - .76	.84	.70 <sup>c</sup>
	KK02 (3)	3.28 (.70)	54 - 61 (57)	.63 - .68	.74	.62 <sup>c</sup>
	KK03 (3)	3.83 (.74)	69 - 74 (71)	.70 - .76	.85	.42 <sup>b</sup>
	KK04 (3)	3.46 (.92)	51 - 71 (62)	.62 - .77	.83	.70 <sup>c</sup>
Gesamtskala	KK01 - KK04 (12)	3.66 (.63)		.45 - .68	.90	
Rollenvorbild und professionelles Handeln	PH01 (3)	3.70 (.86)	65 - 71 (68)	.63 - .76	.83	.80 <sup>c</sup>
	PH02 (3)	4.28 (.69)	77 - 89 (82)	.63 - .67	.81	.69 <sup>c</sup>
	PH03 (3)	3.87 (.70)	66 - 76 (72)	.61 - .76	.83	.64 <sup>c</sup>
Gesamtskala	PH01 - PH03 (9)	3.94 (.58)		.45 - .69	.88	
Systembezogenes Lehren und Lernen	SL01 (3)	3.05 (.88)	40 - 62 (53)	.48 - .65	.74	.55 <sup>c</sup>
	SL02_3)	2.33 (.97)	18 - 48 (32)	.47 - .71	.75	.75 <sup>c</sup>
	SL03_3)	2.63 (.90)	22 - 47 (38)	.34 - .67	.71	.75 <sup>c</sup>
	SL04_3)	3.44 (.94)	51 - 66 (59)	.64 - .72	.81	.60 <sup>a</sup>
Gesamtskala	SL01 - SL04 (12)	2.82 (.74)		.28 - .60	.87	
Reflexion und Weiterentwicklung der eigenen Lehrpraxis	RW01 (4)	3.43 (.86)	47 - 71 (61)	.60 - .70	.83	.76 <sup>c</sup>
	RW02 (3)	2.72 (1.01)	43 (43)	.81 - .91	.93	.73 <sup>c</sup>
	RW03 (3)	2.81 (1.02)	25 - 62 (45)	.46 - .64	.74	.69 <sup>c</sup>
Gesamtskala	RW01 - RW03 (10)	3.03 (.79)		.32 - .85	.88	

<sup>1</sup> In dieser Spalte sind die niedrigsten und die höchsten Schwierigkeitsindices (%) der einzelnen Subskalen aufgeführt. Die Angabe in Klammern bezeichnet den Mittelwert der Schwierigkeitsindices der Items der jeweiligen Subskala.

<sup>2</sup> part-whole korrigierte Trennschärfen

<sup>3</sup> Pearson-Korrelationen: a = <.05, b = <.01, c = p <.001

<sup>4</sup> Globalitems wurden ab WS 2020/2021 nicht mehr erfasst.

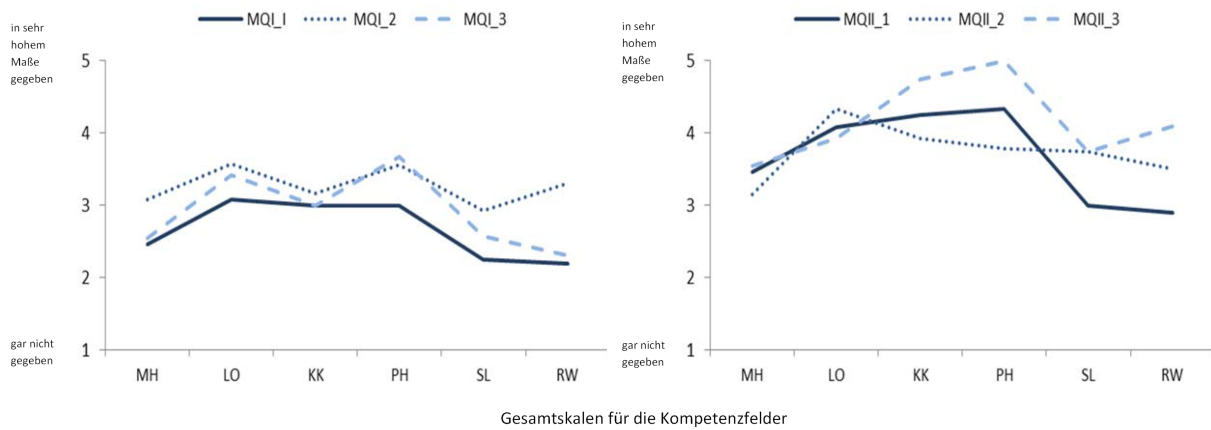
<sup>5</sup> Die Skala wurde für den Einsatz im WS 20/21 und SoSe 2021 um ein Item ergänzt

<sup>6</sup> Die Analysen wurden mit 13 Items durchgeführt, da für diese von allen Probanden Daten vorlagen.

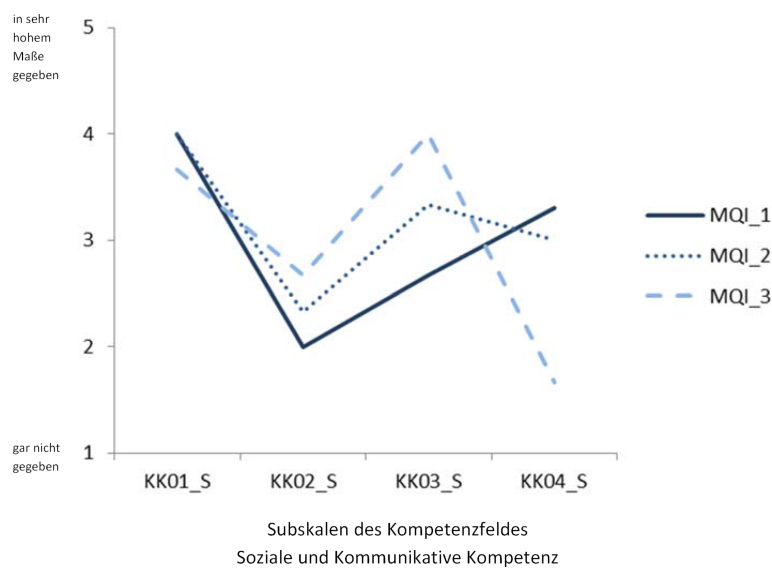
Tabelle 4: Interkorrelation der Gesamtskalen des FKM\_L

	Gesamtskala_MH	Gesamtskala_LO	Gesamtskala_KK	Gesamtskala_PH	Gesamtskala_SL
Gesamtskala_LO	.564 <sup>a</sup>				
Gesamtskala_KK	.600 <sup>a</sup>	.713 <sup>a</sup>			
Gesamtskala_PH	.634 <sup>a</sup>	.492 <sup>a</sup>	.670 <sup>a</sup>		
Gesamtskala_SL	.593 <sup>a</sup>	.317 <sup>a</sup>	.470 <sup>a</sup>	.453 <sup>a</sup>	
Gesamtskala_RW	.693 <sup>a</sup>	.350 <sup>a</sup>	.518 <sup>a</sup>	.570 <sup>a</sup>	.799 <sup>a</sup>

<sup>a</sup> Die Korrelation ist auf dem Niveau von .01 (2-seitig) signifikant; 89 ≤ N ≤ 90



**Abbildung 1: Skalenwerte für die Gesamtskalen der sechs Kompetenzfelder für jeweils drei verschiedene Teilnehmerinnen und Teilnehmer der Kurse MQ I und MQ II**



**Abbildung 2: Skalenwerte für die Subskalen des Kompetenzfeldes Soziale und Kommunikative Kompetenz für drei Teilnehmerinnen und Teilnehmer der MQ I-Kurse**

„1“ Sie orientieren sich bei Rückmeldungen an gängigen Feedbackregeln.

„3“ Sie erkennen und nutzen passende Zeitpunkte für Feedbackgespräche.

„1“ Sie kennen und nutzen verschiedene Feedbackinstrumente.

Diese Angaben können darauf verweisen, dass die Person zwar in ihrer (Lehr-)Praxis bereits Feedbackgespräche führt, bislang jedoch noch keine Kenntnisse von Feedbackregeln und Feedbackinstrumenten hat. Bei einer nach dem Kurs durchgeführten Befragung würde man erwarten, dass diese Items stärker als zutreffend eingeschätzt werden. Eine zuverlässige Interpretation solcher Unterschiede setzt jedoch die Bestimmung geeigneter Konfidenzintervalle voraus.

## Diskussion

Die Analysen zur teststatistischen Überprüfung des FKM\_L erbrachten überwiegend zufriedenstellende Ergebnisse.

Nachfolgend werden die Ergebnisse getrennt für die Subskalen, Gesamtskalen, Globalfragen sowie für die Auswertungsbeispiele diskutiert.

### Subskalen

Für die sechs Kompetenzfelder wurden jeweils drei bis vier Subskalen gebildet, die Teilkompetenzen erfassen. Ausgehend davon, dass Maße der inneren Konsistenz zwischen  $CR-\alpha=.60$  und  $CR-\alpha=.70$  als ausreichend für Gruppenvergleiche [23], [26] angesehen werden und Reliabilitäten von  $CR-\alpha=.80$  als gut gelten [27], liegen die ermittelten Koeffizienten ( $CR-\alpha=.71$  bis  $CR-\alpha=.93$ ) der 22 Subskalen im akzeptablen bis guten Bereich.

Mit der Trennschärfe eines Items wird aufgezeigt, wie gut mit der Beantwortung des Items der zugehörige Skalensummenwert vorhergesagt wird [24]. Alle für die Subskalen ermittelten Trennschärfen liegen mit Werten zwischen .34 und .91 in einem durchweg akzeptablen bis zufriedenstellenden Bereich. Um festzustellen, ob mit den einzelnen Items eine Differenzierung zwischen Personen

mit unterschiedlichen Kompetenzausprägungen möglich ist, wurden für die Items der Subskalen Schwierigkeitsindices berechnet. Zudem wurde für jede Subskala der Mittelwert der Schwierigkeitsindices ihrer Items ermittelt. Nach Döring und Bortz [24], sind extrem schwierige Items (Schwierigkeitsindex <20%) wie auch extrem leichte Items (Schwierigkeitsindex >80%) als wenig informativ hinsichtlich ihrer Fähigkeit zur Differenzierung zwischen den Befragten anzusehen. Die Auswertungen zeigen, dass Letzteres für jeweils eine Subskala der Kompetenzfelder *Lernerorientierung* und *Rollenvorbild und Professionelles Handeln* zutrifft. Eine nähere Betrachtung der Items lässt vermuten, dass dies dadurch zustande kommt, dass die Items vergleichsweise global formuliert sind (z.B. „Sie legen Wert auf gegenseitigen Respekt“) und eventuell zu selbstwertdienlichen Antworten animieren können. Bei drei der Subskalen weisen Items Schwierigkeitsindices zwischen 18% und 25% auf, d.h. dass diese nur von maximal 25 Prozent der Personen als zutreffend eingeschätzt wurden. Zwei dieser Subskalen sind dem Kompetenzfeld *Systembezogenes Lehren und Lernen* zuzuordnen, eine Subskala dem Kompetenzfeld *Reflexion und Weiterentwicklung der eigenen Lehrpraxis*. Die betreffenden Items dieser Subskalen thematisieren in der Regel eine aktive Beteiligung an der Planung und Umsetzung des Curriculums. Die Teilnehmerstruktur der medizindidaktischen Kurse ist allerdings heterogen. Die Teilnehmerinnen und Teilnehmer kommen aus verschiedenen Fächern, sind in unterschiedlichen Positionen und Funktionen tätig und haben mehr oder weniger lange Erfahrungen in der Lehre sammeln können. Damit ist nachvollziehbar, dass sie die oben beschriebenen Items je nach Hintergrund in unterschiedlichem Maße als zutreffend einschätzen können. Die Subskalen mit den zu leichten bzw. zu schwierigen Items können dennoch zur Reflexion der jeweiligen Lehrkompetenzen als didaktisches Mittel eingesetzt werden. Dennoch sollten weitere Analysen durchgeführt werden, um die Items dieser Subskalen eventuell zu ergänzen und/oder umzuformulieren. Ergänzende kognitive Pretests können wertvolle Hinweise liefern, wie diese Items umformuliert werden könnten.

## Gesamtskalen (Kernkompetenzen)

Für jedes Kompetenzfeld des FKM\_L wurden die Items der Subskalen zu Gesamtskalen zusammengefasst. Diese Gesamtskalen weisen durchgehend gute innere Konsistenzen auf, was nachvollziehbar ist, da eine Verlängerung von Skalen mit zusätzlichen Items i.d.R. zu höheren Reliabilitäten führt [23].

Der FKM\_L wurde auf der Basis des Modells *Kernkompetenzen für Lehrkräfte in der Medizin* entwickelt. Bereits bei der Entwicklung dieses Modells wurde festgestellt, dass einige Teilkompetenzen und Lernziele verschiedenen Kompetenzfeldern zuordenbar waren [2]. Von daher lassen sich die moderaten bis hohen Interkorrelationen der Gesamtskalen gut erklären (siehe Tabelle 4). Die aus den Korrelationen zu ermittelnden gemeinsamen Varianzanteile ( $r^2$ ) der Gesamtskalen liegen zwischen .10 und .64.

13 von 15 dieser gemeinsamen Varianzanteile sind <.50, so dass durchaus davon auszugehen ist, dass die Gesamtskalen unterschiedliche Facetten der Lehrkompetenz erfassen. Die sehr hohe Interkorrelation zwischen den Gesamtskalen *Reflexion und Weiterentwicklung der eigenen Lehrpraxis* (RW) und *Systembezogenes Lehren und Lernen* (SL) gibt jedoch Anlass zu überlegen, inwieweit hier eine Zusammenlegung oder Kürzungen sinnvoll sind.

## Globalitems

Die meisten Globalitems korrelieren zwar erwartungsgemäß signifikant, jedoch nicht immer substantiell hoch ( $r > .50$ ) mit den ihnen zugehörigen Subskalen (siehe Tabelle 3). Zudem korrelieren Globalitems eines jeweiligen Kompetenzfeldes häufig auch mit den Subskalen der anderen Kompetenzfelder signifikant und in vergleichbarer Höhe. Dies legt nahe, dass die Globalitems eine geringe diskriminante Validität aufweisen, so dass eine Diagnostik auf Ebene der Globalitems zurzeit nicht zu empfehlen ist. Aus diesem Grund, aber auch um die Beantwortungszeit zu verkürzen und mit dem Aufwand möglicherweise verbundene negative Effekte zu vermeiden, wurde auf die zusätzliche Erfassung dieser Items ab WS 2020/2021 verzichtet.

## Auswertungsbeispiele

Mit den exemplarischen Einzelauswertungen sind inter- und intraindividuelle Unterschiede in den Kompetenzausprägungen erkennbar. Die nächsten Schritte sehen die Ermittlung von Vergleichswerten der jeweiligen Bezugsgruppen vor. Hierzu ist jedoch eine breitere Datenbasis erforderlich.

## Feedback zum Fragebogen

Die Rückmeldungen der Teilnehmerinnen und Teilnehmer der verschiedenen Kurse bezogen sich überwiegend auf den Umfang des Fragebogens, d.h. auf die Anzahl der Items und die Länge der Formulierung einzelner Items. Verschiedentlich wurde angeregt, die Bearbeitungszeit zu erhöhen, da diese für die detaillierte Kompetenzerfassung nicht ausreichend sei. Darüber hinaus wurde die Verwendung von Begriffen kritisiert, die nicht ohne weitere Vorinformation verständlich sind. Auch die Art der Darbietung (v.a. Zuordnung von Items zu Kompetenzen vs. willkürliche Anordnung; Aufteilung auf Bildschirmseiten) wurde angesprochen. Gewünscht wurden kurze Zusatzinformationen in Form von Randnotizen zur Lesart bestimmter Items, um die Einschätzungen verlässlicher zu machen.

## Validität

Bislang wurden lediglich Schritte unternommen, um die Inhaltsvalidität des FKM\_L zu sichern, indem der Entwicklung das KLM-Modells zugrunde gelegt wurde und die Autorinnen und der Autor die Formulierung und Zuord-



nung der Items theoretisch-argumentativ vornehmen [24], [25]. Eine eingehende Überprüfung der Konstrukt- und der Kriteriumsvalidität des Verfahrens steht jedoch noch aus.

## Fazit

Zusammenfassend betrachtet, kann der FKM\_L anhand der beschriebenen Ergebnisse trotz leichter Einschränkungen bereits jetzt zur Individual- und Gruppendiagnostik eingesetzt werden: Lehrende können anhand ihrer FKM\_L-Profile entscheiden, ob sie an der einen oder anderen Stelle ihre erfassten Kompetenzausprägungen überprüfen bzw. optimieren möchten, und Faculty Development-Anbieter können den FKM\_L als Screening-Instrument verwenden, um die Bedarfe der Teilnehmerinnen und Teilnehmer sowie Lücken im Qualifikationsangebot zu ermitteln. Darüber hinaus können die Erfassung von Kompetenzen und deren Ergebnisse im Rahmen medizindidaktischer Kurse als didaktisches Mittel eingesetzt werden, beispielsweise zur Bewusstmachung und Reflexion von individuellen Lehrkompetenzen und deren Entwicklungsbedarf.

## Förderung

Die Studie wurde im Rahmen des BMBF-geförderten Verbundprojekts MERLIN (Medical Education Research – Lehrforschung im Netz BW) der Medizinischen Fakultäten Freiburg, Heidelberg, Mannheim, Ulm und Tübingen unter Federführung der Standorte Freiburg und Tübingen durchgeführt. Förderzeichen TÜ: 01PL17011A, FR: 01PL17011B

## Interessenkonflikt

Die Autor\*innen erklären, dass sie keinen Interessenkonflikt im Zusammenhang mit diesem Artikel haben.

## Literatur

- Bromme R. Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In: Weinert FE, editor. *Enzyklopädie der Psychologie. Psychologie des Unterrichts und der Schule*. Göttingen: Hogrefe; 1997. p.177-212.
- Görlitz A, Ebert T, Bauer D, Grasl M, Hofer M, Lammerding-Köppel M, Fabry G. Core Competencies for Medical Teachers (KLM) – A Position Paper of the GMA Committee on Personal and Organizational Development in Teaching. *GMS Z Med Ausbild*. 2015;32(2):Doc23. DOI: 10.3205/zma000965
- Lipowsky F. Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In: Allemann-Ghionda C, Terhart E, editors. *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern: Ausbildung und Beruf*. Weinheim: Beltz; 2006. p.47-70.
- Srinivasan M, T Li ST, Meyers FJ, Pratt DD, Collins JB, Braddock C, Hilty DM. "Teaching as a Competency": competencies for medical educators. *Acad Med*. 2011;86(10):1211-1220. DOI: 10.1097/ACM.0b013e31822c5b9a 13
- Braun E, Athanassios G, Pollerhof K, Schwabe U. Wie lassen sich kommunikative Kompetenzen messen? Konzeption einer kompetenzorientierten Prüfung kommunikativer Fähigkeiten von Studierende. *Beitr Hochschulforsch*. 2018;40(3):34-55. Zugänglich unter/available from: <https://www.fachportal-paedagogik.de/literatur/vollanzeige.html?Fid=1148973>
- Braun E, Gusy B, Leidner B, Hannover B. Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*. 2008;54(1):30-43. DOI: 10.1026/0012-1924.54.1.30
- Giesler M, Forster J, Biller S, Fabry G. Entwicklung eines Fragebogens zur Erfassung von Kompetenzen in der Medizin: Ergebnisse zur Reliabilität und Validität. *GMS Z Med Ausbild*. 2011;28(2):Doc31. DOI: 10.3205/zma000743
- Gollwitzer M, Jäger RS. *Evaluation – Workshop*. Weinheim: Beltz PVU; 2007.
- Jurecka A. Introduction of the Computer-Based Assessment of Competencies. In: Hartig J, Klieme E, Leutner D, editors. *Assessment of Competencies in Educational Contexts*. Bern: Hogrefe; 2008. p.193-213.
- Hartig J. Psychometric models for the assessment of competencies. In: Hartig J, Klieme E, Leutner D, editors. *Assessment of Competencies in Educational Contexts*. Bern: Hogrefe; 2008. p.69-90.
- Wass V, Vleuten van der C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357(9260):945-949. DOI: 10.1016/S0140-6736(00)04221-5
- Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med*. 2005;80(10 suppl):S46-54. DOI: 10.1097/00001888-200510001-00015
- Peterson LN, Eva KW, Rusticus SA. The readiness for Clerkship Survey: Can Self-assessment Data be used to evaluate Program Effectiveness? *Acad Med*. 2012;87(10):1355-1360. DOI: 10.1097/ACM.0b013e3182676c76
- Colthart I, Bagnell G, Evans A, Allbutt H, Haig A, Illing J, Mckinstry B. The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. *Med Teach*. 2008;30(2):124-145. DOI: 10.1080/01421590701881699
- Jünger J, Schellberg D, Nikendei C. Subjektive Kompetenzeinschätzung von Studierenden und ihre Leistung im OSCE. *GMSZ Med Ausbild*. 2006;23(3):Doc51. Zugänglich unter/available from: <https://www.egms.de/static/de/journals/zma/2006-23/zma000270.shtml>
- Sieverding M. Frauen unterschätzen sich: Selbstbeurteilungs-Biases in einer simulierten Bewerbungssituation. *Z Sozialpsychol*. 2003;34(3):147-160. DOI: 10.1024//0044-3514.34.3.147
- León SP, Vallejo AP, Nelson JB. Variability in the accuracy of self-assessments among low, moderate, and high performing students in university education. *Pract Ass Res Eval*. 2021;26:Article 16. DOI: 10.7275/6q91-az58
- Bußenius L, Harendza S, van den Bussche H, Selch S. Final-year medical students' self-assessment of facets of competence for beginning residents. *BMC Med Educ*. 2022;22(1):82. DOI: 10.1186/s12909-021-03039-2
- Ehrlinger J, Dunning D. How chronic selfviews influence (and potentially mislead) estimates of performance. *J Pers Soc Psychol*. 2003;84(1): 5-17. DOI: 10.1037/0022-3514.84.1.5

20. Regehr G, Hodges B, Tiberius R, Lofchy J. Measuring self-assessment skills: An innovative relative ranking model. *Acad Med.* 1996;71(10 Suppl):852-854. DOI: 10.1097/00001888-199610000-00043
21. Schulz von Thun F. *Miteinander reden 1: Störungen und Klärungen. Allgemeine Psychologie der zwischenmenschlichen Kommunikation.* 53. Auflage. Reinbek (bei Hamburg): Rowohlt-TB; 1981.
22. Griewatz J, Simon M, Lammerding-Koeppel M. Competency-based teacher training: A systematic revision of a proven programme in medical didactics. *GMS J Med Educ.* 2017;34(4):Doc44. DOI: 10.3205/zma001121
23. Lienert GA. *Testaufbau und Testanalyse.* 2. Auflage. Weinheim: Beltz; 1967.
24. Döring N, Bortz J. *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften.* 5. Auflage. Heideberg: Springer; 2016. DOI: 10.1007/978-3-642-41089-5
25. Bühner M. *Einführung in die Test- und Fragebogenkonstruktion.* 3. akt. Auflage. München: Pearson Studium; 2011.
26. Rammstedt B. Zur Bestimmung der Güte von Multi-Item-Skalen: Eine Einführung (ZUMA How-to-Reihe Nr. 12). Mannheim: ZUMA; 2004.
27. Nunnally JC, Bernstein IH. *Psychometric Theory.* 3. Ed. New York: McGraw-Hill; 1994.

**Korrespondenzadresse:**

Dr. Marianne Giesler  
 Freiburg, Deutschland  
 Dr\_M\_Giesler@t-online.de

**Bitte zitieren als**

Giesler M, Lammerding-Köppel M, Griewatz J. Development of a diagnostic instrument for the assessment of teaching competencies in medicine (FKM\_L): First results of the test statistical verification. *GMS J Med Educ.* 2022;39(5):Doc53. DOI: 10.3205/zma001574, URN: urn:nbn:de:0183-zma001574

**Artikel online frei zugänglich unter**

<https://doi.org/10.3205/zma001574>

**Eingereicht:** 05.01.2022

**Überarbeitet:** 28.07.2022

**Angenommen:** 04.08.2022

**Veröffentlicht:** 15.11.2022

**Copyright**

©2022 Giesler et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.