

Catering to the patient – development, validation and psychometric properties of an innovative assessment instrument

Abstract

Introduction: It has been shown that communication skills acquired during undergraduate medical education are of great importance. Hence, many countries require teaching communication as part of their medical curricula. To assess students' learning progress, "Catering to the Patient", as an aspect of showing empathy, should be evaluated. Since there was no description of a validated instrument fitting for this purpose, one had to be developed. To describe its process of development and its psychometric properties were the aims of this study.

Methods: Based on the Calgary-Cambridge Observation Guide (CCOG), items describing catering to the patient were selected and modified. Cognitive pretest interviews were conducted to check understandability. Therefore, 7 raters assessed 1 video each ($R=7$, $V=1$). In a following pilot study ($R=3$, $V=10$) first psychometric properties were evaluated and necessary corrections in the preliminary evaluation form were carried out before the final evaluation form was used to assess students' ability to cater to the patient and psychometric properties were described in detail ($R=2$, $V=35$).

Results: The final assessment instrument, "catering to the patient – Marburg evaluation form", contains 11 checklist items and two global ratings (items 12 and 13). In the final evaluation the inter-rater reliability (IRR) ranged from 0 to 0.562, the median was $r=0.305$. Concerning item 13 (a global rating), 88.6% of the videos were scored with the maximum difference of one point. The internal consistency was very high (Cronbach's α : $\alpha=0.937$ and $\alpha=0.962$), and the correlation between the checklist items and the global rating was high (Pearson's r : $r=0.856$ and $r=0.898$).

Discussion: The assessment instrument "catering to the patient" is suitable for giving feedback and for using it in formative examinations. Its use for summative examinations can be considered. Further examinations should evaluate if a three-point Likert scale could reach higher values and if item 13 can be used as a stand-alone item.

Keywords: assessment, communication, empathy, communication curriculum, medical education

1. Introduction

It is estimated that doctors conduct 200,000 conversations with patients throughout their careers [1]. Just by conducting a thorough medical history, 70% of all diseases can be diagnosed correctly, and with the aid of a physical examination, 90% [2]. These facts alone show the great importance of communication in medicine.

Additionally, the change in the physician-patient relationship in the second half of the 20th century, from paternalism to a patient centered approach, has brought a change to the expectations of physicians' communication [3].

Physicians can improve their communication abilities. When physicians participated in a one month [4] or 2.5

day module, their communications skills improved [5]. These abilities can already be acquired throughout undergraduate education [6]. It has been shown that the communication skills acquired during undergraduate education are of great importance [7]. Hence, many countries require teaching communication as part of their medical curricula. For example, in Canada, it is integrated into The CanMEDS Physician Competency Framework [8] and in Germany in the licensing act for physicians ("Approbationsordnung für Ärzte") [https://www.gesetze-im-internet.de/_appro_2002/].

Often, assessment instruments are used for teaching and evaluating students' communication skills. There are many instruments to rate these skills [9], [10], [11], [12].

Miriam Urff¹

Antonio Krüger²

Steffen Ruchholtz³

Eva Christina Stibane⁴

1 Universitätsklinik Freiburg, Klinik für Psychiatrie, Psychotherapie und Psychosomatik im Kindes- und Jugendalter, Freiburg, Germany

2 Asklepios Klinik Lich, Unfallchirurgie, Orthopädie, Wirbelsäulen- und Kindertraumatologie, Lich, Germany

3 UKGM Standort Marburg, Zentrum für Orthopädie und Unfallchirurgie, Marburg, Germany

4 Philipps-Universität Marburg, Zentrum für Medizinische Lehre, Marburg, Germany

Most of them focus on the different steps in a conversation. Norfolk et al. view empathy as a key aspect for developing rapport [13]. Therefore, it is a core competence of doctor–patient communication. Additionally it is known that empathy decreases during medical school and postgraduate medical education, which makes it very important to enhance it in medical education [14].

In the context of teaching students communication skills and empathy the responsible for the communication curriculum of the skills-lab in Marburg discussed, that it would be valuable to evaluate and ensuing teaching students' abilities in catering to the patient. This is not the same as empathy but shows a great overlap. Catering to the patient puts the focus on the activities and behavior while empathy is a basic attitude.

The Calgary-Cambridge Observation Guide (CCOG) is an assessment instrument developed by S. Kurtz and J. Silverman to teach and evaluate students' communicational abilities. It comprises in the long version 73 items in seven categories: initiating the session, gathering information, providing structure, building relationship, explanation and planning, closing the session, options in explanation and planning. The short version is structured correspondingly. The CCOG is a well described [1], [15], [16] and widely used instrument [17], [18], [19], [20] that addresses many aspects of doctor–patient communication, including but not only aspects of catering to the patient. Since the assessment instrument “catering to the patient” should be easy to use and focus on aspects of catering to the patient the new evaluation form was developed based on the CCOG.

Since a suitable assessment instrument for evaluating these aspect of catering to the patient could not be found and even Julie M. Schirmer, author of “Assessing Communication Competence: A Review of Current Tools” [10] did not know about a fitting instrument, one had to be developed.

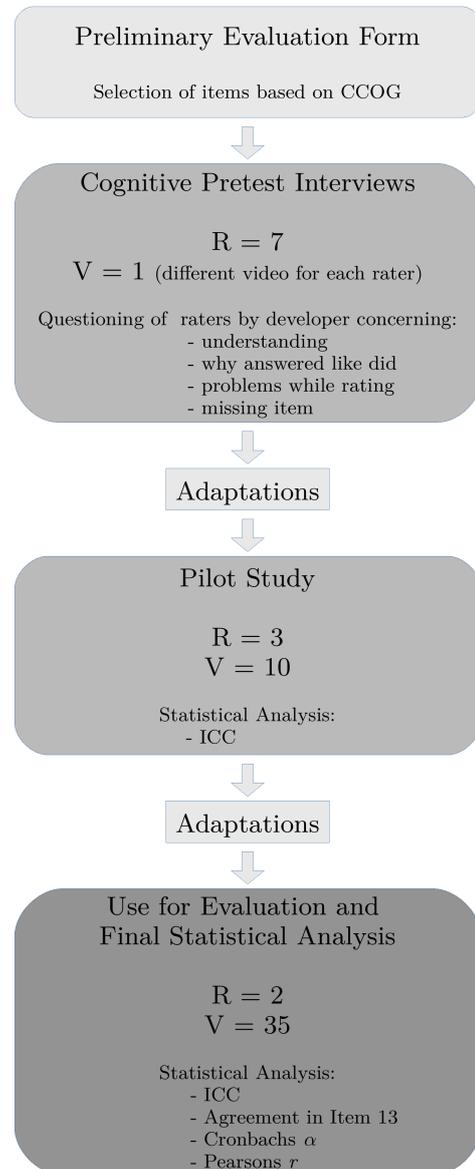
The aim of this study was to develop an assessment instrument that describes the aspects of catering to the patient as a core competence of doctor-patient communication. It should be easy and time efficient to use and contain good psychometric properties. Therefore, the newly developed instrument should be examined concerning these properties.

2. Methods

The assessment instrument “catering to the patient” was constructed and reviewed in a German version. On purpose of publication the instrument was translated into English.

As it can be seen in figure 1 based on the CCOG a preliminary evaluation form was assembled. Following cognitive pretest interviews were conducted. For that purpose seven raters watched each one different video of a doctor-patient encounter ($R=7$, $V=1$), in the following they were asked about the wording and idea of the items. Consecutive necessary adaptations were carried out and in a

next step it was tested in a pilot study: three raters rated each ten same videos ($R=3$, $V=10$). Again, adaptations were made before it was used to assess students' ability in catering to the patient. Therefore all 35 videos of patient-centered doctor's advices were assessed by two raters ($R=2$, $V=35$). The obtained data were applied to describe its psychometric properties in detail. For each phase different raters were used.



R: number of raters

V: number of videos assessed by each rater

Figure 1: Steps of developing and evaluating the new assessment instrument “catering to the patient”

The ethics committee of Philipps-University Marburg approved the study (Az:22/13), and all participants gave informed consent.

2.1. Assembling of the preliminary evaluation form

The assembling of the assessment instrument as well as the decisions on how to transform items were mainly conducted by an experienced peer-teacher, a skills-lab leader and an expert of development of questionnaires. Doubtful decisions were discussed with the skills-lab team responsible for the communicational curriculum.

Based on the long version of the CCOG [15], consisting of 73 items, all items concerning the doctor–patient relationship and showing empathy were selected, resulting in 24 selected items (items 6, 10, 12, 14, 15, 17, 18, 23, 32, 42, 43, 45, 46, 64, 66, 71). Then, all items that describe a behavior that is not essential for catering to the patient were excluded (items 6, 17, 18, 24, 25, 26, 28, 30, 31, 46, 64). This was done to not rate students worse, who could not show these behaviors due to the observed situation. Some items concerning similar activities were combined (items 12 and 45, 14 and 42). Item 24 was excluded because students in observed situations during the lesson rarely take notes or use a computer. Item 45 had to be split and modified to avoid a double-barreled item. Some items were selected after minor changes (items 10, 23, 29, 32, 42, 66, 71) without changing their meaning; for example in item 23 the explanatory notes were put in brackets. The items then were modified and worded for an observation instrument to fit the format: The student shows a described behavior. To finalize the questionnaire, a global rating (*“Altogether the student has catered very well to the patient’s needs.”*) was added.

The preliminary evaluation form consisted of 11 checklist items and was concluded by a global rating out of rater perspective.

Concerning the scoring, it was decided to use a Likert scale, and additionally short descriptions for each point (“fully applies” to “does not apply at all”). Researching the literature regarding the number of points used when scoring the CCOG showed that five-point Likert scales [12], [21] and three-point Likert scales [1], [22] were used. A five-point Likert scale was chosen to better distinguish between different levels of catering to the patient.

2.2. Cognitive pretest interviews

Qualitative cognitive pretest interviews were carried out to identify any problems concerning the wording of the items and answering categories [23]. Seven students, working as peer teachers in the skills lab, were asked to watch and consecutively rate each one different videotaped interview of a student and a simulated patient (SP). Afterwards, they were asked whether they understood the items, why they chose to answer as they did, whether they had problems understanding any of the items, and whether they felt an aspect of catering to the patient was missing. It was also tested whether a category “not applicable” should be given. For reasons of practicability, pre-existing videos were used, although it was not possible

to answer two items: these two items concerned examining the patient and concluding the interview, the interview was stopped after eight minutes and did not include a physical examination. Nevertheless, the videos were suitable because the themes of the discussion were taboo subjects (partnership and sexuality), prompting catering to the patient and developing rapport to a high degree. The topic of the doctor-patient encounter was a patient presenting with worsening of the symptoms of an autoimmune disease. The doctor can learn from the patient that he is not taking his prescribed cortisone because of the side effects and their consequences on his life and his feelings. Initially the videos were recorded for a study concerning difficult topics in doctor-patient communication.

2.3. Pilot study

The same 10 videos that were used before were now watched by 3 different raters using the assessment tool. Two raters were students who work in the skills-lab, and the third rater was an expert who developed the assessment instrument.

Intra class correlation (ICC) was used to assess inter-rater reliability (IRR) because it can be used to compare the ratings of three or more raters. r as coefficient for ICC is confined to an interval from -1 to +1. Since the values of the measurements of reliability are confined to the interval from 0 to +1, negative values show a reliability of 0 [24]. Values of $0.25 < r < 0.5$ describe a moderate correlation, and $r > 0.5$ describes a strong correlation [25]. All items with a reliability of $r < 0.25$ were modified after the pilot study.

2.4. Use of the assessment instrument and final statistical analysis

The last step was to use the evaluation form to assess students’ ability in catering to the patient and to evaluate the assessment instrument on a larger sample. Therefore, two raters rated 35 videos. One rater was an anesthesiology resident, and the other was a resident in child and adolescent psychiatry and psychotherapy. The videos were recorded in connection with a study that evaluated the effectiveness of a course for medical students about difficult doctor–patient encounters.

Each of the videos showed the interaction between a student and an SP. The students’ task was each time to discover the reason for the patient’s admission to the hospital (including taking a patient’s history and physical examination); and informing the patient about follow-up procedures. The setting for the appointment was a patient with symptoms suggesting a bronchial carcinoma. He would like to discuss his fear of pending death with the doctor. The underlying educational topic of discussion is catering to the patient. The duration of the videos is about 20 minutes.

IRR was assessed again, to detect whether the changes in the assessment instrument made after the pilot study

were effective. For comparability, ICC was used again. Additionally it was compared how often the residents gave the same rating for item 13. The percentage of agreement as well as the percentage of difference of one, two, three or four points were analyzed. Internal consistency was evaluated using Cronbach's α [26]. α may vary between $-\infty$ and $+1$; ideally it reaches values around 0.9 [27]. To assess whether the global rating of item 13 reflects items 1 to 11, an index was created. This index was determined for each video by calculating the mean value of items 1 to 11. Then the correlation between the index and item 13 was assessed. (For calculating the index, item 12 was not integrated because it is a global rating from the patient's perspective).

3. Results

The results of the cognitive pretest interviews, the following adaptations, the pilot study, and the subsequent changes in two items led to the assessment instrument "catering to the patient – Marburg evaluation form" shown in figure 2. Its psychometric properties were evaluated in a final statistical analysis.

3.1. Results of the cognitive pretest interviews

As a result of the pretest interviews, some minor changes in the wording of the existing items were made. Furthermore, an additional item was integrated, regarding a global rating from the patient's perspective. It was included as item 12, after the items addressing different behaviors and before the global rating from the observer's perspective.

The pretest showed that if the answering category "not applicable" was given, items that could be judged often were not answered. Therefore, it was decided to not include that option.

3.2. Results of the pilot study

Within the pilot study, ICC was calculated to evaluate IRR. The results are shown in table 1.

Items 8 and 11 could not be evaluated because no physical examination took place and the conversation was stopped after eight minutes. Items 5, 6, 12 and 13 showed moderate correlation between raters and items 1, 4, 7, 9 and 10 showed strong correlation. Due to their good reliability, these items were not changed. Items 2 and 3 showed a reliability of 0, therefore changes had to be made within the items. It was observed that these two items were double-barreled, which might have caused their low reliability.

Hence, item 2, which originally read: "The student formulates questions and explanations that are exact and simple to understand (avoids using medical terminology or explains them well)" was changed to "The student formulates questions and explanations that are simple to understand (avoids using medical terminology or ex-

plains them well)". The wording of the item was changed to assess only whether the questions were simply formulated. In doing so, the patient's understanding is taken into consideration.

Item 3 also had to be modified. The item originally read: "The student shows that he emphatically understands the patient's feelings and situation (e.g., by verbalizing, paraphrasing)" was changed to "The student shows that he emphatically understands the patient's feelings and situation". It was referred that the examples complicated the rating, therefore they were excluded.

3.3. Description of the assessment instrument, "catering to the patient"

The assessment instrument used for the final assessment is shown in figure 2. It contains 13 items, of which 11 describe different behaviors and two are global ratings from the patient's as well as the observer's perspective. They are rated with the aid of a five-point Likert scale, labeled with 1=fully applies, 2=rather applies, 3=partly applies, 4=rather does not apply and 5=does not apply at all.

The 11 checklist items contain

1. listening attentively,
2. formulating questions and explanations that are simple to understand,
3. showing emphatically understanding,
4. understanding and responding to verbal indications,
5. understanding and responding to nonverbal indications,
6. displaying appropriate nonverbal behavior,
7. reacting in a sensitive manner,
8. explaining during the physical examination what one is doing,
9. catering to the patient's attitude concerning his situation,
10. bearing in mind the social and cultural background, and
11. ending the conversation in agreement.

The two global ratings are out of patient's perspective (The patient feels he is in good care and understood) and out of rater's perspective (Altogether the student has catered very well to the patient's needs).

3.4. Final statistical analysis

The analysis of the psychometric properties in the final assessment included an evaluation of the IRR, the agreement between the raters in item 13, internal consistency, as well as the correlation of the index with item 13.

The results of the ICC (see table 2), expressing the IRR, range from $r=-0.153$ to $r=0.562$. The median is $r=0.305$. It can be seen that the reliability of items 2, 6 and 10 is high. Items 1, 4, 5, 7, 11 and 13 show moderate reliability. Item 3 shows low reliability and items 8, 9 and 12 show a reliability of $r=0$.

“Catering towards the Patient” – Marburg Evaluation Form

Instruction: It is necessary to observe the entire conversation before making an overall evaluation. However, please acquaint yourself with the questions before observing the conversation. You may use the reverse side for notes.

Please rate using the following scale:

1 2 3 4 5
fully applies rather applies partly applies rather does not apply does not apply at all

		1	2	3	4	5
1	The student listens attentively while the patient speaks, allowing for pauses.					
2	The student formulates questions and explanations that are simple to understand (avoids using medical terminology or explains them well).					
3	The student shows that he ¹ empathically understands the patients feelings and situation.					
4	The student understands and responds to verbal indications (e.g. wishes, questions, apprehensions and fears).					
5	The student understands and responds to nonverbal indications (e.g. wishes, questions, apprehensions and fears).					
6	The student displays appropriate nonverbal behavior towards the patient (e.g. eye contact, facial expressions, body language, movement, speech, tone of voice).					
7	The student inquires, or picks up on, delicate matters and reacts in a sensitive manner .					
8	The student explains during the physical examination what he is doing .					
9	The student caters towards the patient’s attitude concerning his situation (e.g. opinion, worries, expectations).					
10	The student bears in mind the social and cultural background of the patient and adjusts his conversation accordingly (e.g. choice of words, explanations).					
11	The student and the patient end their conversation in agreement.					
12	The patient feels he is in good care and understood .					
13	Altogether the student has catered very well towards the patient’s needs.					

¹For clarity purposes male pronouns are used. We wish to explicitly state that this is inclusive of all genders.

Figure 2: Assessment instrument “Catering to the Patient”

Table 1: Intra class correlation (ICC) and following inter-rater reliability (IRR) of pilot study

Item	1	2	3	4	5	6
ICC	0.548	-0.421	-0.328	0.538	0.455	0.362
IRR	0.548	0	0	0.538	0.455	0.362

Item	7	8	9	10	11	12	13
ICC	0.738	#	0.627	0.561	#	0.387	0.471
IRR	0.738	#	0.627	0.561	#	0.387	0.471

This item cannot be evaluated on the basis of the videos.

Table 3 shows the results of agreement in item 13. All videos scored with a 1 by rater 1 are displayed in row 1. The columns indicate how rater 2 rated these videos. So five videos were scored with a 1 by both raters; two videos were scored with a 1 by rater 2 and a 2 by rater 1. It can

be seen that 45.7% (n=16) of the videos were scored the same by both raters and there is only a one point difference in the ratings for 42.9% (n=15) of them. Those that differ by two points constitute 8.6% (n=2), and those with a three-point difference, 2.9% (n=1).

Table 2: Intra class correlation (ICC) and following inter-rater reliability (IRR) of final assessment

Item	1	2	3	4	5	6	7
ICC	0.435	0.562	0.205	0.305	0.279	0.556	0.259
IRR	0.435	0.562	0.205	0.305	0.279	0.556	0.259

Item	8	9	10	11	12	13
ICC	-0.153	-0.059	0.501	0.335	-0.120	0.453
IRR	0	0	0.501	0.335	0	0.453

Table 3: Agreement in item 13 between raters

		Rating of rater 1				
		1	2	3	4	5
Rating of rater 2	1	5	2	2	1	0
	2	5*	8	3	0	0
	3	1	3	3	2	0
	4	0	0	0	0	0
	5	0	0	0	0	0

* concerning item 13 there are 5 videos that were rated with 1 by rater 1 and 2 by rater 2, etc.

Internal consistency of the assessment instrument is evaluated by using Conbach’s α . With $\alpha=0.937$ for rater 1 and $\alpha=0.962$ for rater 2 it shows excellent values and therefore a high interrelatedness and internal consistency of the items.

Table 4 shows the means of the index and item 13 as well as Pearson’s r to describe the correlation. The correlation between the index and item 13 for rater 1, using Pearson’s r , is $r=0.856$, and for rater 2, $r=0.898$. The significance for both values is $p<0.001$. This indicates a high correlation between items 1 to 11 and item 13.

Table 4: Correlation of index and item 13

	Mean of index ¹	Mean of item 13	Pearson’s r
Rater 1	1.86	2.09	0.856 **
Rater 2	1.88	1.97	0.898 **

¹ mean values of items 1 to 11 create index; ** $p < 0.001$

4. Discussion

The CCOG is a widely used instrument for teaching [18] and measuring [17], [19], [20] communication. The short version of the CCOG was translated into German and examined by a German research group. The test-retest reliability and validity are good, therefore it is suitable for evaluating medical students’ communication abilities [16].

The purpose of this study was to develop an assessment instrument to measure students’ ability to cater to the patient. The instrument should be easy to use and time-efficient. Furthermore, its psychometric properties should be good.

4.1. Choosing the answering category

Simmenroth-Nayda et al. used a three-point Likert scale in their pretest, but changed it to a five-point Likert scale because the pretest revealed problems with ratings, which led to that change [16]. On the other hand, a three-point

scale might have been able to reach higher values for IRR [28]. In the future, it needs to be decided whether a higher IRR is necessary, in which case a three-point scale could be used. Otherwise, if a more differentiated rating is important (e.g., for giving feedback), then a five-point scale should be used. As described in 2.1, it was decided to use a five-point Likert scale.

4.2. Statistical analysis

The statistical analysis suggests that the assessment instrument is feasible for teaching and formative examinations. For summative examinations it can be recommended, keeping in mind that some of the results are only moderate.

The results of the pilot study showed that the reliability of items 2 and 3 is 0. When the ICC was calculated in the final assessment, these items showed an ICC-coefficient of $r=0.562$ for item 2 and $r=0.205$ for item 3. These results suggest that the changes were, especially for item 2, effective. Both calculations of IRR showed moderate values for the medians: in the pilot study, the median was $m=0.471$, and in the final assessment, $m=0.305$. It is possible that a three-point Likert scale might have led to higher values [28]. A German evaluation of the short version of the CCOG observed similar results to the ones discussed in this paper [16]. They reported IRR-coefficients between 0.05 and 0.57.

Another possibility to improve the reliability might be through rater training [29]. Although rater training varies, and in some cases the raters only receive information via e-mail [30], it was decided not to conduct any rater training. The decision was made because training, even in a digital way, is time consuming, which conflicts with the aim of a simple and time-efficient assessment instrument. Additionally, often no rater training occurs when instruments, such as this one, are used outside the context of a study. Therefore, the antecedent study must also be done without any rater training.

Concerning item 13, 45.7% of the videos were scored the same by both raters, while 42.9% of the videos were scored with a one-point difference. This results in 88.6% of the videos being scored with no more than a one-point difference. Whether this is acceptable depends on the purpose of its usage. For formative examinations and feedback it is suitable. Before using in a summative evaluation, one should consider whether the score should be used alone or combined with scores of other assessment instruments.

For rating complex skills like communication, global ratings are better suited than checklists [30], [31]. In this

evaluation it could be shown that the correlation between the checklist items (items 1 to 11) and the global rating (item 13) is very high ($r_1=0.856$, $r_2=0.898$). Additionally, excellent internal consistency was reached ($\alpha_1=0.937$, $\alpha_2=0.926$). Based on this data it could be deduced that item 13 by itself is sufficient to rate the catering to the patient. On the other hand, the checklist items probably teach the rater about the criteria for evaluation before the global rating is given. Thus, whether item 13 by itself is suitable for summative examination needs further evaluation. For formative examination the items can be used to give more detailed feedback.

4.3. Critical limitations

Focusing on the critical limitations first the small samples are to be pointed out. Already during the pilot study a greater amount of data would have helped to improve the items. Especially it would have been advisable to conduct a second pilot study to figure out if the changes made were effective. In the final statistical analysis it would have been advantageous to integrate the data of a larger number of raters to get better values for IRR. If the number of videos had been higher, statistical analysis would have been sounder.

Another limitation is the aim to measure the construct of “Catering to the Patient”. The decision to assess this concept arose from practical needs, although it is seen as good scientific practice to select the aim of measuring on the basis of theoretical considerations or literature research. Additionally there was no conceptual preliminary study, so the construct is not well defined. Therefore, a clear differentiation between empathy, developing rapport and catering to the patient was not made. Otherwise, it is an opportunity to use a tool that is based on practical needs.

In the study out of practical reasons only videos showing conversations between students and SPs were rated. It might be that the observation of direct encounters as well as conversations between real patients and students would show different results. It is reckoned that the differences between SPs and real patients is small if SPs act well. The difference between videotaped encounters and live encounters is hard to estimate since there might be differences in what the rater can observe e.g. the gestures and facial expression as well as the atmosphere between the two parties. In this study the raters were not allowed to re-watch scenes or videos because this would probably make a big difference compared to live encounters. Additionally when in use outside of a study there would probably not be enough time to watch videos or even parts of them several times.

The aim was to establish an assessment instrument that is simple to use. Nevertheless it is necessary to take some minutes before observing the encounter to get acquainted with the items. Since it is fundamental to watch the entire conversation before rating, again time is needed after the encounter to conduct the rating. (That is why) To make

it as time-efficient as possible, the data arising from this study were produced without any rater training

Another limitation is about the language. In this study the validation was conducted using the German version. Only for this article it was translated into English. Additionally it is suggested to take different steps in the translation process e.g. a back translation and various reviews [32]. In this case the evaluation form was translated by the author and a native speaker in English, both with excellent knowledge in the other language, without any further translation process.

4.4. Future prospects

Further studies could analyze if the patient’s or SP’s evaluation of the students’ ability to cater to the patient correlates with the assessment of a not involved observer. For this item 12 (global rating out of patient’s perspective) can be used.

This assessment instrument has a wide range for future use. As described above it can be deployed in training as well as in examination, bearing in mind that its use in summative evaluations has to be critically checked. In this study it was only utilized to evaluate videotaped situations but it is imaginable to be used in live encounters as well. It is estimated that it can be applied in all levels of the medical education. It was tested in two undergraduate settings (pilot study and final statistical analysis) but the use in a postgraduate setting is conceivable as well.

Probably most people think about employing the instrument in subjects focusing on the conversation with the patient but it can be used in other subjects as well. The topics of the encounter recorded in the videos of the final statistical analysis were in connection with teaching surgical students. Even though it could be used in all kind of situations it will probably be most advantageous in training or assessing of demanding situations.

5. Conclusion

The psychometric properties of the assessment instrument “Catering to the Patient – Marburg Evaluation Form” suggest that it definitely can be used for teaching and formative examination. Its use can be considered for summative examination. Nevertheless, it should be evaluated whether a three-point Likert scale could reach higher values of IRR. Furthermore, it would be interesting to examine whether item 13 as a stand-alone item could be used, since this was considered to be controversial.

Acknowledgements

We thank all the students and raters who contributed to the success of this study. Particularly we want to thank Lorianna Könniker for her help in translating the assessment instrument. We also want to acknowledge the fin-

ancial support of the Department of General Medicine of Philipps-University Marburg, who made it possible for the raters to receive an expense allowance; the “Förderverein Maris e.V.,” who paid for the costs of the proofreading service; and the skills-lab “Maris” itself for covering the expenditure on the peer teachers’ wages.

Profiles

Name of school: Philipps University Marburg

Study program/occupation: Medicine

Number of students per year and/or per semester: 350 p.a.

Has a longitudinal curriculum covering communication been implemented? Yes

At which semester levels are communicative and social competencies taught? 1, 2, 5, 6, 7, 8, 9, (PJ)

Which teaching formats are used? Courses, practical courses, simulation with simulated patients

During which semesters are communicative and social competencies tested (formative, pass/fail, graded)? Formative: 2 and 9, summative with marks 5 and 6

Which assessment formats are used? Roleplay, video feedback, OSCE with simulated patients

Who (e.g. hospital, institution) is in charge of development and implementation? Interdisciplinary skills lab in cooperation with clinical departments

Current professional roles of the authors

- Dr. med. Miriam Urff studied medicine in Marburg and is assistant doctor at the University Medical Center Freiburg, Clinic for Psychiatry, Psychotherapy and Psychosomatics in Childhood and Adolescence. Three years she was student tutor in the Marburg interdisciplinary skills lab and conducted a project for doctor-patient’s communication in surgery for her dissertation.
- Prof. Dr. med. Antonio Krüger is the clinic’s head physician for trauma surgery, orthopedics, spine and pediatric trauma at the Asklepios Clinic in Lich. He was a senior physician at the university clinic in Marburg for many years, and he was lecturer and agent for students in the practical year. He is convinced that teaching interdisciplinary – for example in skills labs – can improve learning.
- Prof. Dr. med. Steffen Ruchholtz is Director of the Center for Orthopedics and Trauma Surgery at the university hospital Marburg. As a representative for the Trauma Network Committee of the German Society for Trauma Surgery he is also responsible for the implementation of a concept to optimize the clinical quality of treatment for seriously injured patients in Germany.
- Dr. phil. Tina Stibane heads the Center for Medical Teaching and Learning at the Philipps University of

Marburg. She coordinates the skills lab Maris, supervises various teaching-learning projects and is active part of the curriculum committee. She also conducts the medical-didactic training for lecturers and tutors in the medical department.

Data

Data for this article are available from the Dryad Digital Repository: <https://www.doi.org/10.5061/dryad.qjq2bvqcw> [33]

Competing interests

The authors declare that they have no competing interests.

References

1. Kurtz S, Silverman J, Draper J. Teaching and Learning Communication Skills in Medicine. 2nd ed. Oxford, San Francisco: Radcliffe Publishing; 2005.
2. Schweickhardt A, Fritzsche K. Kursbuch ärztliche Kommunikation: Grundlagen und Fallbeispiele aus Klinik und Praxis; mit 15 Tabellen. 2nd ed. Köln: Deutscher Ärzte-Verlag; 2009.
3. Berian JR, Ko CY, Angelos P. Surgical Professionalism: The Inspiring Surgeon of the Modern Era. *Ann Surg.* 2016;263(3):428-429. DOI: 10.1097/SLA.0000000000001425
4. Smith RC, Lyles JS, Mettler J, Stoffelmayr BE, Van Egeren LF, Marshall AA, Gardiner JC, Maduschke KM, Stanley JM, Osborn GG, Shebroe V, Greenbaum RB. The effectiveness of intensive training for residents in interviewing. A randomized, controlled study. *Ann Intern Med.* 1998;128(2):118-126. DOI: 10.7326/0003-4819-128-2-199801150-00008
5. Levinson W, Roter D. The effects of two continuing medical education programs on communication skills of practicing primary care physicians. *J Gen Intern Med.* 1993;8(6):318-324. DOI: 10.1007/BF02600146
6. Yedidia MJ, Gillespie CC, Kachur E, Schwartz MD, Ockene J, Chepaitis AE, Snyder CW, Lazare A, Lipkin Jr M. Effect of communications training on medical student performance. *JAMA.* 2003;290(9):1157-1165. DOI: 10.1001/jama.290.9.1157
7. Kiessling C, Dieterich A, Fabry G, Hölzer H, Langewitz W, Mühlhans I, Pruskil S, Scheffer S, Schubert S. Basler Consensus Statement "Kommunikative und soziale Kompetenzen im Medizinstudium": Ein Positionspapier des GMA-Ausschusses Kommunikative und soziale Kompetenzen. *GMS Z Med Ausbild.* 2008 Mai;25(2):Doc83. Zugänglich unter/available from: <https://www.egms.de/static/de/journals/zma/2008-25/zma000567.shtml>
8. Frank JR, editor. The CanMEDS 2005 physician competency framework Better standards. Better physicians. Better care. Ottawa: The Royal College of Physicians and Surgeons of Canada; 2005.
9. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. *Acad Med.* 2001;76(4):390-393. DOI: 10.1097/00001888-200104000-00021

10. Schirmer JM, Mauksch L, Lang F, Marvel MK, Zoppi K, Epstein RM, Brock D, Pryzbylski M. Assessing communication competence: a review of current tools. *Fam Med*. 2005;37(3):184-192.
11. Sennekamp M, Gilbert K, Gerlach FM, Guethlin C. Development and validation of the "FrOCK": Frankfurt observer communication checklist. *Z Evid Fortbild Qual Gesundheitswes*. 2012;106(8):595-601. DOI: 10.1016/j.zefq.2012.07.018
12. Simmenroth-Nayda A, Weiss C, Fischer T, Himmel W. Do communication training programs improve students' communication skills? - a follow-up study. *BMC Res Notes*. 2012;5:486. DOI: 10.1186/1756-0500-5-486
13. Norfolk T, Birdi K, Walsh D. The role of empathy in establishing rapport in the consultation: a new model. *Med Educ*. 2007;41(7):690-697. DOI: 10.1111/j.1365-2923.2007.02789.x
14. Hojat M. *Empathy in Health Professions Education and Patient Care*. 1st ed. Berlin, Heidelberg: Springer; 2016. DOI: 10.1007/978-3-319-27625-0
15. Silverman J, Kurtz S, Draper J. *Skills for Communicating With Patients*. 3rd ed. London New York: Radcliffe Publishing; 2013.
16. Simmenroth-Nayda A, Heinemann S, Nolte C, Fischer T, Himmel W. Psychometric properties of the Calgary Cambridge guides to assess communication skills of undergraduate medical students. *Int J Med Educ*. 2014;5:212-218. DOI: 10.5116/ijme.5454.c665
17. Bosse HM, Schultz JH, Nickel M, Lutz T, Möltner A, Jünger J, Huwendiek S, Nikendei C. The effect of using standardized patients or peer role play on ratings of undergraduate communication training: a randomized controlled trial. *Patient Educ Couns*. 2012;87(3):300-306. DOI: 10.1016/j.pec.2011.10.007
18. Choudhary A, Gupta V. Teaching communications skills to medical students: Introducing the fine art of medical practice. *Int J Appl Basic Med Res*. 2015;5(Suppl 1):S41-44. DOI: 10.4103/2229-516X.162273
19. Nguyen LN, Tardioli K, Roberts M, Watterson J. Development and incorporation of hybrid simulation OSCE into in-training examinations to assess multiple CanMEDS competencies in urologic trainees. *Can Urol Assoc J*. 2015;9(1-2):32-36. DOI: 10.5489/cuaj.2366
20. Roh H, Park KH, Jeon YJ, Park SG, Lee J. Medical students' agenda-setting abilities during medical interviews. *Korean J Med Educ*. 2015;27(2):77-86. DOI: 10.3946/kjme.2015.27.2.77
21. Ammentorp J, Thomsen JL, Jarbøl DE, Holst R, Øvrehus AL, Kofoed PE. Comparison of the medical students' perceived self-efficacy and the evaluation of the observers and patients. *BMC Med Educ*. 2013;13:49. DOI: 10.1186/1472-6920-13-49
22. Harasym PH, Woloschuk W, Cuning L. Undesired variance due to examiner stringency/leniency effect in communication skill scores assessed in OSCEs. *Adv Health Sci Educ Theory Pract*. 2008;13(5):617-632. DOI: 10.1007/s10459-007-9068-0
23. Schnell R, Hill PB, Elke E. *Methoden der empirischen Sozialforschung*. 10th ed. München, Oldenbourg: De Gruyter Oldenbourg; 2013.
24. Wirtz M, Caspar F. *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen, Bern, Toronto, Seattle: Hogrefe - Verlag für Psychologie; 2002.
25. Kühnel SM, Krebs D. *Statistik für die Sozialwissenschaften: Grundlagen, Methoden, Anwendungen*. 6th ed. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag; 2012.
26. Schmitt N. Uses and Abuses of Coefficient Alpha. *Psychol Ass*. 1996;8(4):350-353. DOI: 10.1037/1040-3590.8.4.350
27. Kline P. *The Handbook of Psychological Testing*. London, New York: Routledge; 1993.
28. Nolte C. *Validität und Reliabilität eines Instruments zur Messung der Qualität der Kommunikation und seine Eignung im studentischen Unterricht*. Göttingen: Georg-August-Universität zu Göttingen; 2014.
29. Crossley J, Humphris G, Jolly B. Assessing health professionals. *Med Educ*. 2002;36(9):800-804. DOI: 10.1046/j.1365-2923.2002.01294.x
30. Scheffer S. *Validierung des "Berliner Global Rating" (BGR) - ein Instrument zur Prüfung kommunikativer Kompetenzen Medizinstudierender im Rahmen klinischpraktischer Prüfungen (OSCE)*. Berlin: Charité - Universität Berlin; 2009.
31. Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ*. 2004;38(2):199-203. DOI: 10.1111/j.1365-2923.2004.01755.x
32. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P; ISPOR Task Force for Translation and Cultural Adaptation. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*. 2005;8(2):94-104. DOI: 10.1111/j.1524-4733.2005.04054.x
33. Urff M, Krüger A, Ruchholtz S, Stibane T. "Catering to the patient" - psychometric properties of an innovative assessment instrument. *Dryad Digital Repository*. 2021. DOI: 10.5061/dryad.qjq2bvqow

Corresponding author:

Dr. Eva Christina Stibane
 Philipps-Universität Marburg, Zentrum für Medizinische
 Lehre, Conradstr. 7, D-35043 Marburg, Germany, Phone:
 +49 (0)6421/28-25053
 stibane@staff.uni-marburg.de

Please cite as

Urff M, Krüger A, Ruchholtz S, Stibane EC. *Catering to the patient – development, validation and psychometric properties of an innovative assessment instrument*. *GMS J Med Educ*. 2021;38(3):Doc69. DOI: 10.3205/zma001465, URN: urn:nbn:de:0183-zma0014650

This article is freely available from

<https://www.egms.de/en/journals/zma/2021-38/zma001465.shtml>

Received: 2020-03-31

Revised: 2020-07-14

Accepted: 2020-09-21

Published: 2021-03-15

Copyright

©2021 Urff 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/>.

Das Eingehen auf Patienten – Entwicklung, Validierung and psychometrische Eigenschaften eines neuen Erhebungsinstruments

Zusammenfassung

Einleitung: Es konnte gezeigt werden, dass kommunikative Fähigkeiten, die während des Medizinstudiums erworben werden, von großer Wichtigkeit sind. Daher wird in vielen Ländern Kommunikation, als Teil des Curriculums des Medizinstudiums, zu lehren gefordert.

Um den Lernfortschritt von Studierenden zu beurteilen, sollte das „Eingehen auf den Patienten“ als ein Aspekt von Empathie evaluiert werden. Da es keine Beschreibung eines validierten Instruments gab, das für diesen Zweck genutzt werden konnte, musste ein solches entwickelt werden. Das Ziel dieser Abhandlung ist, den Entwicklungsprozess des Instruments und seine psychometrischen Eigenschaften zu beschreiben.

Methoden: Basierend auf dem Calgary-Cambridge Observation Guide (CCOG), wurden Items, die das Eingehen auf den Patienten beschreiben, ausgewählt und modifiziert. Kognitive Pretest-Interviews wurden durchgeführt um die Verständlichkeit zu überprüfen. Dafür bewerteten 7 Rater jeweils 1 Video (R=7, V=1). In einer darauffolgenden Pilotstudie (R=3, V=10) wurden das erste Mal die psychometrischen Eigenschaften evaluiert und notwendige Verbesserungen am vorläufigen Bewertungsbogen durchgeführt, bevor der finale Bewertungsbogen genutzt wurde, um die Fähigkeiten von Studierenden bezüglich des Eingehens auf den Patienten zu bewerten und die psychometrischen Eigenschaften detailliert zu beschreiben (R=2, V=35).

Ergebnisse: Der finale Bewertungsbogen „Eingehen auf den Patienten – Marburger Bewertungsbogen“ beinhaltet 11 Checklisten-Items und zwei Globalbewertungen (Items 12 und 13). In der abschließenden Evaluation wies die Interrater-Reliabilität (IRR) Werte zwischen 0 und 0,562 auf, der Median lag bei $r=0,305$. Bezüglich Item 13 (eine Globalbewertung) wurden 88,6% der Videos mit maximal einem Punkt Differenz bewertet. Die Interne Konsistenz war sehr hoch (Cronbach's α : $\alpha=0,937$ und $\alpha=0,962$); die Korrelation zwischen den Checklist-Items und der Globalbewertung war hoch (Pearson's r : $r=0,856$ und $r=0,898$).

Diskussion: Der Bewertungsbogen „Eingehen auf den Patienten“ ist geeignet Feedback zu geben sowie zur Nutzung in formativen Prüfungen. Der Einsatz in summativen Prüfungen kann erwogen werden. Weitere Untersuchungen sollten evaluieren, ob mit einer dreistufigen Likert-Skala höhere Kennwerte erreicht werden könnten und ob Item 13 als alleiniges Item genutzt werden könnte.

Schlüsselwörter: Instrument, Kommunikation, Empathie, Kommunikationscurriculum, Medizinische Ausbildung

1. Einleitung

Geschätzt führen Ärzt*innen rund 200.000 Patientengespräche während Ihrer Berufstätigkeit durch [1]. Schon durch eine gründliche Anamneseerhebung können 70% aller Erkrankungen korrekt diagnostiziert werden, mit ei-

ner zusätzlichen körperlichen Untersuchung 90% [2]. Allein diese Daten zeigen die große Bedeutung der Kommunikation in der Medizin.

Darüber hinaus hat die Veränderung von einer paternalistischen zu einer Patienten-zentrierten Arzt-Patientenbeziehung in der zweiten Hälfte des 20. Jahrhunderts zu veränderten Erwartungen an die ärztliche Kommunikation geführt [3].

Miriam Urff¹

Antonio Krüger²

Steffen Ruchholtz³

Eva Christina Stibane⁴

1 Universitätsklinik Freiburg, Klinik für Psychiatrie, Psychotherapie und Psychosomatik im Kindes- und Jugendalter, Freiburg, Deutschland

2 Asklepios Klinik Lich, Unfallchirurgie, Orthopädie, Wirbelsäulen- und Kindertraumatologie, Lich, Deutschland

3 UKGM Standort Marburg, Zentrum für Orthopädie und Unfallchirurgie, Marburg, Deutschland

4 Philipps-Universität Marburg, Zentrum für Medizinische Lehre, Marburg, Deutschland

Ärzt*innen können ihre kommunikativen Fähigkeiten verbessern. Wenn Ärzt*innen, an einem einmonatigen [4] oder zweieinhalb-tägigen-Kurs teilnahmen, verbesserten sich ihre kommunikativen Fähigkeiten [5]. Diese Fähigkeiten können schon während des Medizinstudiums erworben werden [6]. Es konnte sogar gezeigt werden, dass die während des Studiums erworbenen kommunikativen Fertigkeiten von enormer Bedeutung sind [7]. Viele Länder haben in ihre Medizin-Curricula die Lehre von Kommunikation aufgenommen. In Kanada zum Beispiel ist sie Bestandteil des Kompetenz-Bezugsrahmens für Ärzt*innen CanMEDS [8] und in Deutschland der Ärztlichen Approbationsordnung [https://www.gesetze-im-internet.de/_appro_2002/].

Häufig werden Fragebögen für den Unterricht und die Evaluation studentischer Kommunikationsfähigkeiten eingesetzt. Es gibt viele Instrumente, um diese Fähigkeiten zu messen [9], [10], [11], [12]. Die meisten von ihnen fokussieren auf verschiedene Gesprächsphasen. Norfolk et al. sehen Empathie als Schlüssel-Aspekt für einen gelingenden Beziehungsaufbau an [13]. Deshalb sei sie eine Kernkompetenz in der Arzt-Patienten-Kommunikation. Gleichzeitig ist bekannt, dass Empathie während des Studiums und der Weiterbildung abnimmt, was ihre Förderung in der medizinischen Ausbildung umso wichtiger macht [14].

Im Kontext des Unterrichts kommunikativer Fertigkeiten und Empathie diskutierten die Verantwortlichen für das Kommunikationscurriculum im Marburger Skills-Lab, dass es sinnvoll wäre, die Fähigkeit der Studierenden, auf Patienten einzugehen zu evaluieren und zu fördern. Diese ist nicht gleichzusetzen mit Empathie, zeigt aber eine große Überschneidung. Das Eingehen auf Patienten nimmt die Aktivität und das Verhalten mehr in den Blick, die Empathie eher die Grundeinstellung.

Der Calgary-Cambridge Observation Guide (CCOG) ist ein von S. Kurtz und J. Silverman entwickeltes Messinstrument für die Lehre und Evaluation von Kommunikationsfähigkeiten von Studierenden. In der Langversion umfasst es 73 Items in sieben Kategorien: Gesprächseröffnung, Informationssammlung, Strukturierung, Beziehungsgestaltung, Erklären und Planen, Gesprächsabschluss, zusätzliche Möglichkeiten bezüglich Erklären und Planen. Die Kurzversion ist gleichermaßen strukturiert. Der CCOG ist ein ausführlich beschriebenes [1], [15], [16] und häufig genutztes Instrument [17], [18], [19], [20], das viele Aspekte der Arzt-Patienten-Kommunikation adressiert, einschließlich, aber nicht nur des Aspektes des „Eingehens auf Patienten“.

Da ein passendes Messinstrument für die Evaluation des Aspektes „Eingehen auf Patienten“ nicht gefunden wurde und auch Julie M. Schirmer, Autorin von „Assessing Communication Competence: A Review of Current Tools“ [10] keines kannte, sollte ein Instrument entwickelt werden.

Das Ziel der vorliegenden Arbeit war es, ein Messinstrument zu entwickeln, das den Aspekt des Eingehens auf Patienten als Kernkompetenz der Arzt-Patienten-Kommunikation beschreibt. Es sollte einfach und zeiteffizient in der Anwendung sein sowie gute psychometrische Eigenschaften besitzen. Dafür musste das neu entwickelte Instrument hinsichtlich dieser Eigenschaften untersucht werden.

Das Messinstrument „Eingehen auf Patienten“ wurde in einer deutschen Fassung entwickelt und untersucht. Für die Veröffentlichung wurde das Instrument ins Englische übersetzt.

2. Methoden

Das Messinstrument „Eingehen auf Patienten“ wurde in einer deutschen Fassung entwickelt und untersucht. Für die Veröffentlichung wurde das Instrument ins Englische übersetzt.

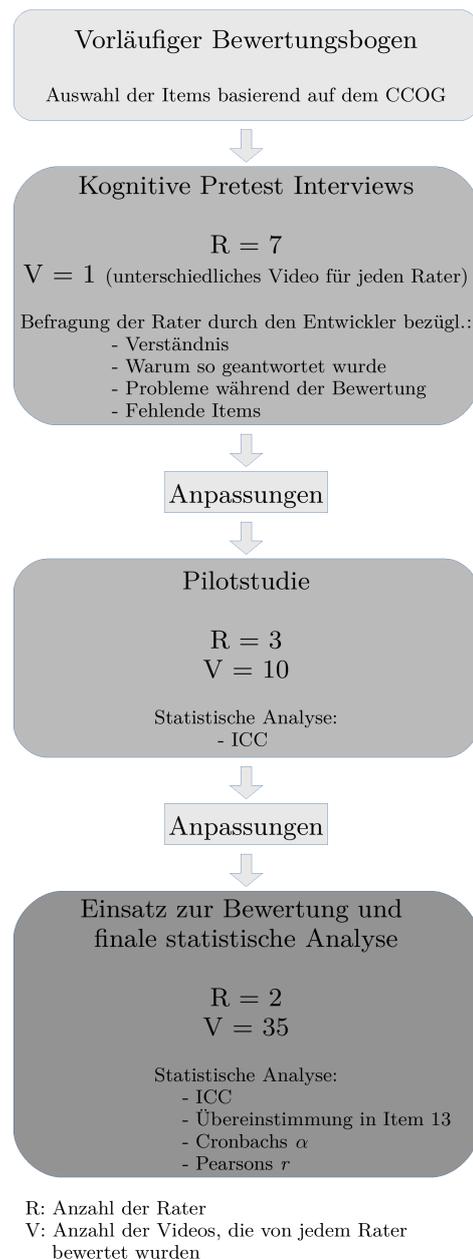


Abbildung 1: Schritte während der Entwicklung und Überprüfung des neuen Bewertungsbogens „Eingehen auf den Patienten“

Wie in Abbildung 1 zu sehen ist, wurde, basierend auf dem CCOG, ein erster Bogen als Entwurfsfassung zusam-

mengestellt. Daraufhin wurden kognitive Interviews durchgeführt. Für diesen Zweck sahen sich sieben Personen jeweils ein unterschiedliches Video an (R=7, V=1) und wurden zur Formulierung und zum Verständnis der Items befragt. Anschließend wurden notwendige Anpassungen vorgenommen und im nächsten Schritt das Instrument in einer Pilot-Studie untersucht: drei Rater sahen sich jeweils die selben zehn Videos an (R=3, V=10). Erneut wurde das Instrument überarbeitet, bevor es für die Erfassung der Fähigkeit der Studierenden, auf Patienten einzugehen, eingesetzt wurde. Hierfür sahen zwei Rater jeweils die gleichen 35 Videos von ärztlichen Beratungsgesprächen an und bewerteten diese mithilfe des Instruments. Die gewonnenen Daten wurden für die detaillierte Beschreibung der psychometrischen Eigenschaften ausgewertet. In jeder Phase wurden verschiedene Personen als Rater eingesetzt.

Die Ethik-Kommission der Philipps-Universität genehmigte die Studie (AZ:22/13) und alle Teilnehmer*innen gaben ihr Einverständnis.

2.1. Der Aufbau des Bewertungsbogens in der Entwurfssfassung

Der Aufbau des Bewertungsbogens sowie die Entscheidung einzelne Items aufzunehmen, wurde in einem Diskussionsprozess zwischen einer erfahrenen studentischen Tutorin, der Skills Lab-Leiterin und einem Methodiker und Experten für Fragebogenentwicklung erarbeitet. Umstrittene Entscheidungen wurden mit der für das Kommunikationscurriculum verantwortlichen Skills Lab-Teamleitung diskutiert.

Basierend auf der 73 Items umfassenden Langversion des CCOG [15] wurden zunächst alle 24 Items eingeschlossen, die die Arzt-Patienten-Beziehung und das Zeigen von Empathie betreffen (Items 6, 10, 12, 14, 15, 17, 18, 23 - 32, 42, 43, 45, 46, 64, 66, 71). Daran anschließend wurden alle Items, die ein Verhalten beschreiben, das nicht wesentlich für das Eingehen auf Patienten ist, ausgeschlossen (Items 6, 17, 18, 24, 25, 26, 28, 30, 31, 46, 64). Dies, um Studierende nicht schlechter zu bewerten, die aufgrund der beobachteten Situation, die entsprechenden Verhaltensweisen nicht zeigen könnten. Items, die sich auf sehr ähnliche Verhaltensweisen bezogen, wurden verbunden (Items 12 und 45, 14 und 42). Item 24 wurde ausgeschlossen, da Studierende in Beobachtungssituationen im Unterricht kaum Notizen machen oder einen Computer benutzen. Item 45 wurde in zwei Items geteilt und modifiziert, um doppelte Stimuli innerhalb eines Items zu vermeiden. Einige Items wurden nach kleinen Veränderungen eingeschlossen (Items 10, 23, 29, 32, 42, 66, 71); zum Beispiel wurde bei Item 23 die Erläuterung in Klammern gesetzt. Die Items wurden für ein Fragebogen-Format formuliert: Der Student zeigt ein beschriebenes Verhalten. Zu guter letzt wurde eine Globalbewertung angefügt ("Student geht insgesamt sehr gut auf den Patienten ein").

Die Entwurfssfassung des Bewertungsbogens bestand aus 11 Items und einer abschließenden Globalbewertung aus der Beobachterperspektive.

Bezüglich der Bewertung wurde entschieden, eine Likert-Skala und zusätzlich kurze Beschreibungen für jeden Punkt („trifft voll und ganz zu“ bis „trifft gar nicht zu“) zu nutzen. Eine Literaturrecherche bezüglich der Anzahl der Stufen zur Bewertung des CCOG zeigte, dass Fünfstufige [12], [21] und Dreistufige [1], [22] genutzt worden waren. Eine fünfstufige Likert-Skala wurde gewählt, um differenzierter zwischen den verschiedenen Niveaus des Eingehens auf den Patienten unterscheiden zu können.

2.2. Kognitive Pretest-Interviews

Qualitative kognitive Interviews für den Pretest wurden durchgeführt um Probleme in der Formulierung der Items oder der Antwortkategorien zu identifizieren [23]. Sieben Studierende, die als Tutor*innen im Skills Lab arbeiteten wurden gebeten, jeweils ein anderes auf Video aufgenommenes Arzt-Patientengespräch von Studierenden und Schauspielern (SP) anzusehen und mittels Bewertungsbogen zu bewerten. Sie wurden gefragt, wie sie die Items verstanden, warum sie ihre Antwort so wählten, ob sie Probleme im Verständnis von Items hatten und ob ihnen ein Aspekt des Eingehens auf den Patienten fehlte. Auch wurde gefragt, ob eine Antwortkategorie „nicht beantwortbar“ möglich sein sollte. Aus Praktikabilitätsgründen wurden bereits existierende Videos verwendet, obwohl es anhand dieser nicht möglich war, zwei Items zu bewerten: diese betrafen die körperliche Untersuchung und den Gesprächsabschluss, die Aufnahme endete nach 8 Minuten und beinhaltete keine körperliche Untersuchung. Dennoch waren die Videos geeignet, denn die Themen des Gesprächs behandelten Tabu-Themen (Partnerschaft und Sexualität) und erforderten ein hohes Maß an Eingehen auf den Patienten bzw. die Patientin und Beziehungsaufbau. Das Thema der ärztlichen Konsultation war die Verschlechterung der Symptomatik einer Autoimmunerkrankung. Die Ärzt*in kann – bei einfühlsamer Gesprächsführung – vom Patienten erfahren, dass dieser das verordnete Cortison-Präparat nicht einnimmt, wegen der damit verbundenen Nebenwirkungen und den Auswirkungen auf sein Leben und seine Gefühlswelt. Ursprünglich waren die Videos aufgenommen worden für eine Studie, die schwierige Themen der Arzt-Patienten-Kommunikation behandelte.

2.3. Die Pilot-Studie

Die bereits zuvor genutzten 10 Videos wurden nun von drei verschiedenen Personen angesehen und mit Hilfe des Instruments bewertet. Zwei von ihnen waren Tutor*innen des Skills Labs, der dritte ein Experte, der das Instrument mitentwickelt hatte.

Die Intraklassen-Korrelation (ICC), die den Grad an Übereinstimmung zwischen zwei und mehr Ratern abbildet wurde genutzt, um die Interrater-Reliabilität (IRR) zu messen. r , als Koeffizient der ICC, ist begrenzt auf ein

Intervall von -1 bis +1. Da die Reliabilität durch Werte zwischen 0 und +1 angegeben wird, bedeuten negative Werte der ICC eine Reliabilität von 0 [24]. Dabei stehen Werte von $0,25 < r < 0,5$ für eine moderate Übereinstimmung und $r > 0,5$ für eine hohe Übereinstimmung [25]. Alle Items mit einer Reliabilität von $r < 0,25$ wurden im Anschluss an die Pilot-Studie überarbeitet.

2.4. Der Einsatz des Instruments und seine statistische Analyse

Im letzten Schritt wurde das Instrument eingesetzt um die Fähigkeit der Studierenden, auf Patienten einzugehen, zu messen und das Instrument anhand einer größeren Stichprobe zu validieren. Zu diesem Zweck bewerteten zwei Rater 35 Videos. Einer war Assistenzarzt der Anästhesiologie, der andere Assistenzarzt der Kinder- und Jugendpsychiatrie und Psychotherapie. Die Videos waren im Zusammenhang mit einer Studie aufgenommen worden, die die Lerneffekte des Unterrichts in herausfordernden ärztlichen Gesprächssituationen untersuchte. Jedes der Videos zeigte die Interaktion zwischen einer/einem Studierenden und einer Simulationsperson. Die Aufgabe für die Studierenden war jedes Mal, den Grund für die Vorstellung im Krankenhaus zu erkennen (einschließlich Anamneseerhebung und körperlicher Untersuchung) und den Patienten bzw. die Patientin über das weitere Vorgehen zu informieren. Der Anlass des Termins waren Symptome, die das Vorhandensein eines Bronchial-Carcinoms nahelegen. Die Patient*innen möchten ihre Todes-Ängste mit der Ärzt*in besprechen. Das zugrundeliegende pädagogische Thema der Gesprächssituation ist das Eingehen auf einen Patienten, seine Bedürfnisse, Äußerungen und Andeutungen. Die Videos sind etwa 20 Minuten lang. Erneut wurde die IRR berechnet um zu ermitteln, ob die Veränderungen des Instruments infolge der Pilot-Studie zu den gewünschten Effekten geführt hatten. Zum Vergleich, wurde die ICC ein weiteres Mal verwendet. Zusätzlich wurde verglichen wie häufig die Assistenzärzte das Item 13 gleich bewerteten. Die prozentuale Übereinstimmung sowie prozentuale Differenz in ein, zwei, drei oder vier Punkten wurde berechnet. Die Interne Konsistenz wurde mit Cronbachs Alpha beschrieben [26]. Alpha kann zwischen $-\infty$ und +1 variieren; idealerweise erreicht es Werte um 0,9 [27]. Um darzustellen, ob die Bewertung des Globalen Bewertungsitems 13 die Items 1 bis 11 repräsentiert, wurde ein Index gebildet. Dieser Index berechnete sich auf der Basis des Mittelwertes jedes Items von 1 bis 11. Dann wurde die Korrelation zwischen diesem Index und dem Item 13 berechnet. (Für die Berechnung des Index wurde das Item 12, das eine Globalbewertungen aus Patientenperspektive ist, nicht einbezogen).

3. Ergebnisse

Die Ergebnisse der kognitiven Interviews, die darauffolgenden Anpassungen, die Pilotstudie und ihre anschlie-

ßenden Änderungen in zwei Items führten zum Bewertungsbogen „Eingehen auf den Patienten – Marburger Bewertungsbogen“, vorgestellt in Abbildung 2. Seine psychometrischen Eigenschaften wurden in einer statistischen Analyse berechnet.

3.1. Ergebnisse der Kognitiven Pretest Interviews

Als Ergebnis der kognitiven Pretest-Interviews wurden geringfügige Veränderungen in der Formulierung der Items vorgenommen. Darüber hinaus wurde ein weiteres Item eingefügt, eine globale Bewertung aus Patientenperspektive. Es wurde als Item 12 hinter den Items, die verschiedene Verhaltensweisen adressierten und vor dem Item der globalen Bewertung aus der Beobachterperspektive, integriert.

Im Pretest zeigte sich, dass Items, die bewertet werden konnten, häufig nicht bewertet wurden, wenn die Antwortkategorie „nicht bewertbar“ zur Wahl stand.

3.2. Ergebnisse der Pilotstudie

In der Pilotstudie wurde die ICC berechnet um die IRR zu bestimmen. Die Ergebnisse finden sich in Tabelle 1.

Die Items 8 und 11 konnten nicht bewertet werden, da keine körperliche Untersuchung stattgefunden hatte und das Gespräch nach 8 Minuten gestoppt worden war. Die Items 5, 6, 12 und 13 zeigten eine moderate Korrelation zwischen den Bewertern, die Items 1, 4, 7, 9 und 10 eine hohe Korrelation. Wegen ihrer guten Reliabilität wurden diese Items nicht mehr verändert. Die Items 2 und 3 zeigten eine Reliabilität von 0, deshalb wurden sie überarbeitet. Diese beiden Items hatten sich je auf zwei Aspekte bezogen, was der Grund für ihre schlechte Reliabilität sein konnte.

Item 2 lautete ursprünglich „Student formuliert Fragen und Kommentare genau und leicht verständlich (vermeidet Fachsprache oder erklärt diese gut)“ und wurde geändert in „Student formuliert Fragen und Kommentare leicht verständlich (vermeidet Fachsprache oder erklärt diese gut)“. Der Wortlaut des Items wurde dahingehend geändert, dass nur bewertet werden sollte, ob Fragen leicht verständlich formuliert worden waren. Damit wurde bereits auf das Verständnis der Patient*in ausreichend Bezug genommen.

Auch das Item 3 musste modifiziert werden. Das Item lautete zuvor: „Student macht deutlich, dass er die Gefühle des Patienten und seine Lage einfühlsam versteht (z.B. durch Verbalisieren, Paraphrasieren)“ und wurde geändert in „Student macht deutlich, dass er die Gefühle des Patienten und seine Lage einfühlsam versteht“. Es war berichtet worden, dass die Beispiele in Klammern die Bewertung eher erschwerten; sie wurden deshalb entfernt.

„Eingehen auf den Patienten“ – Marburger Bewertungsbogen

Name des Studierenden¹: _____

Anleitung: Es ist notwendig das **gesamte Gespräch** zu beobachten, bevor sie eine Gesamtbeurteilung geben. Bitte machen Sie sich jedoch zuvor mit den Fragen vertraut.

Auf der Rückseite dieses Bogens können Sie sich Notizen machen.

Bitte bewerten Sie folgendermaßen:

1 2 3 4 5
trifft voll und ganz zu trifft eher zu teils/teils trifft eher nicht zu trifft gar nicht zu

		1	2	3	4	5
1	Student hört aufmerksam zu und lässt Patient Zeit zu antworten bzw. lässt Pausen im Patientenbericht zu.					
2	Student formuliert Fragen und Kommentare leicht verständlich (vermeidet Fachsprache oder erklärt diese gut).					
3	Student macht deutlich, dass er die Gefühle des Patienten und seine Lage einführend versteht .					
4	Student greift verbale Hinweise auf und geht auf sie ein (z.B. auf Wünsche, Fragen, Unbehagen, Ängste).					
5	Student greift nonverbale Hinweise auf und geht auf sie ein (z.B. auf Wünsche, Fragen, Unbehagen, Ängste).					
6	Student verhält sich nonverbal dem Patienten gegenüber in angemessener Weise (z.B. Blickkontakt, Gesichtsausdruck, Körperhaltung, Bewegung, Sprache, Stimme).					
7	Student spricht heikle Themen an bzw. greift sie auf und geht sensibel damit um .					
8	Student erklärt während der körperlichen Untersuchung, was er tut .					
9	Student bezieht seine Erläuterungen auf die Sichtweise des Patienten (z.B. Meinungen, Sorgen, Erwartungen).					
10	Student berücksichtigt den sozialen und kulturellen Hintergrund des Patienten und passt das Gespräch (z.B. Wortwahl, Erläuterungen) daran an.					
11	Student beendet das Gespräch im Einvernehmen mit dem Patienten.					
12	Patient fühlt sich gut aufgehoben und verstanden .					
13	Student geht insgesamt sehr gut auf den Patienten ein.					

¹Zugunsten der Übersichtlichkeit und Verständlichkeit werden in diesem Fragebogen stets männliche Formen verwendet. Wir weisen ausdrücklich darauf hin, dass damit gleichermaßen Personen jeglichen Geschlechts gemeint sind.

Abbildung 2: Bewertungsbogen „Eingehen auf den Patienten“

Tabelle 1: Intraklassen-Korrelation (ICC) und die daraus folgende Inter-Rater-Reliabilität (IRR) der Pilotstudie

Item	1	2	3	4	5	6
ICC	0,548	-0,421	-0,328	0,538	0,455	0,362
IRR	0,548	0	0	0,538	0,455	0,362

Item	7	8	9	10	11	12	13
ICC	0,738	#	0,627	0,561	#	0,387	0,471
IRR	0,738	#	0,627	0,561	#	0,387	0,471

Dieses Item kann anhand der Videos nicht bewertet werden.

3.3. Beschreibung des Bewertungsbogens „Eingehen auf den Patienten“

Der Bewertungsbogen, der für die endgültige Datenerhebung

genutzt wurde, wird in Abbildung 2 gezeigt. Er beinhaltet 13 Items von denen 11 verschiedene Verhaltensweisen beschreiben und zwei Globalbewertungen aus der Patienten- bzw. Beobachterperspektive. Die Items wurden mittels einer 5-Punkte-Likert-Skala bewertet, die

definiert sind als 1=trifft voll und ganz zu, 2=trifft eher zu, 3= tells/teils, 4= trifft eher nicht zu und 5=trifft gar nicht zu.

Die 11 Items des Bogens beinhalten

1. aufmerksames Zuhören,
2. leicht verständliche Formulierungen,
3. einführendes Verstehen,
4. Aufgreifen und Eingehen auf verbale Hinweise,
5. Aufgreifen und Eingehen auf nonverbale Hinweise,
6. nonverbal angemessenes Verhalten,
7. sensibles Reagieren,
8. Erklärung der Handlungen während der körperlichen Untersuchung,
9. Bezug auf die Sichtweise des Patienten,
10. Berücksichtigung des sozialen und kulturellen Hintergrunds und
11. Beendigung des Gesprächs im Einvernehmen.

Die beiden globalen Bewertungen sind aus der Patientenperspektive („Patient fühlt sich gut aufgehoben und verstanden“) formuliert und aus der Beobachterperspektive („Student geht insgesamt sehr gut auf den Patienten ein“).

3.4. Abschließende statistische Analyse

Die Analyse der psychometrischen Eigenschaften im endgültigen Bewertungsbogen schloss die Berechnung des IRR, die Übereinstimmung der Bewerter in Item 13, die interne Konsistenz sowie die Korrelation des Index mit Item 13 ein.

Die Ergebnisse des ICC (siehe Tabelle 2), ausgedrückt in der IRR liegt zwischen $r=-0.153$ bis $r=0.562$. Der Median liegt bei $r=0.305$. Die Reliabilität der Items 2, 6 und 10 ist hoch. Die Items 1, 4, 5, 7, 11 und 13 zeigen eine moderate Reliabilität. Das Item 3 zeigt eine nur geringe Reliabilität und die Items 8, 9 und 12 zeigen eine Reliabilität von $r=0$.

Tabelle 3 zeigt die Ergebnisse der Übereinstimmung im Item 13. Alle Videos die von Rater 1 mit 1 bewertet wurden, sind in Zeile 1 dargestellt. Die Spalten zeigen, wie Rater 2 diese Videos bewertete. Es wurden also fünf Videos mit einer 1 von beiden Bewertern bewertet. Zwei Videos wurden mit einer 1 von Rater 2 und einer 2 von Rater 1 bewertet. 45,7% (N=16) der Videos wurden von beiden Ratern gleich bewertet und in 42,9% der Bewertungen liegen beide Bewertungen nur einen Punkt auseinander (N=15). Zwei Punkte auseinander liegen 8,6% (N=2), eine drei-Punkte-Differenz weisen 2,9% (N=1) der Videos auf.

Tabelle 3: Übereinstimmung in Item 13 zwischen den Ratern

		Bewertung von Rater 1				
		1	2	3	4	5
Bewertung von Rater 2	1	5	2	2	1	0
	2	5*	8	3	0	0
	3	1	3	3	2	0
	4	0	0	0	0	0
	5	0	0	0	0	0

* Item 13 betreffend gab es 5 Videos, die von Rater 1 mit 1 und von Rater 2 mit 2 bewertet, etc.

Tabelle 4 führt die Mittelwerte des Index und des Items 13 auf sowie Pearson's r, um die Korrelation zu beschreiben. Die Korrelation zwischen dem Index und Item 13 für den Rater 1, Pearson's r berechnend, ist $r=0.856$ und für den Rater 2 $r=0.898$. Die Signifikanz für beide Werte ist $p<0.001$. Dies zeigt eine hohe Korrelation zwischen den Items 1 bis 11 und Item 13.

Table 4: Korrelation des Index mit Item 13

	Mittelwert des Index ¹	Mittelwert von Item 13	Pearson's r
Rater 1	1,86	2,09	0,856 **
Rater 2	1,88	1,97	0,898 **

¹ Mittelwerte der Items 1 bis 11 bilden Index; ** $p < 0,001$

4. Diskussion

Der CCOG ist ein weit verbreitetes Instrument in der Lehre [18] und zum Messen von Kommunikationskompetenz [17], [19], [20]. Die kurze Version des CCOG wurde ins Deutsche übersetzt und von einer deutschen Forschergruppe untersucht. Die Reliabilität und Validität sind gut, weshalb sie geeignet ist, die kommunikativen Fähigkeiten von Studierenden der Medizin zu evaluieren [16].

Der Zweck der vorliegenden Studie war es, einen Bewertungsbogen zu entwickeln, der die Fähigkeit der Studierende auf den Patienten einzugehen misst. Das Instrument sollte einfach zu nutzen und zeitlich effizient sein. Ferner sollten seine psychometrischen Eigenschaften gut sein.

4.1. Wahl der Antwortkategorien

Simmenroth-Nayda et al. nutzten eine dreistufige Likert-Skala in ihrem Pretest, änderten sie dann jedoch in eine fünfstufige Likert-Skala, da der Pretest Probleme bei der Bewertung zeigte, die zu dieser Änderung führte [16]. Auf der anderen Seite wäre es möglich gewesen durch eine dreistufige Likert-Skala höhere Werte der IRR zu erreichen [28]. In Zukunft sollte entschieden werden, ob eine höhere IRR notwendig erscheint und in diesem Fall eine dreistufige Likert-Skala genutzt werden könnte. Wenn allerdings eine differenziertere Bewertung notwendig ist (z.B. um Feedback zu geben), sollte eine fünfstufige Likert-Skala genutzt werden.

Wie in 2.1 beschrieben, wurde sich hier für eine fünfstufige Likert-Skala entschieden.

Tabelle 2: Intraklassen-Korrelation (ICC) und die daraus folgende Interrater-Reliabilität (IRR) der finalen Auswertung

Item	1	2	3	4	5	6	7
ICC	0,435	0,562	0,205	0,305	0,279	0,556	0,259
IRR	0,435	0,562	0,205	0,305	0,279	0,556	0,259

Item	8	9	10	11	12	13
ICC	-0,153	-0,059	0,501	0,335	-0,120	0,453
IRR	0	0	0,501	0,335	0	0,453

4.2. Statistische Analyse

Die statistische Analyse legt nahe, dass der Bewertungsbogen für die Lehre und formative Prüfungen genutzt werden kann. Für summative Prüfungen kann er empfohlen werden, allerdings sollte bedacht werden, dass manche Ergebnisse nur im moderaten Bereich liegen.

Die Ergebnisse der Pilotstudie zeigten, dass die Reliabilität der Items 2 und 3 bei 0 liegt. Bei der Berechnung der ICC in der finalen Bewertung, zeigten diese Items einen ICC-Koeffizienten von $r=0,562$ für Item 2 and $r=0,205$ für Item 3. Diese Ergebnisse legen nahe, dass die Veränderungen, insbesondere für Item 2, effektiv waren. Beide Berechnung der IRR zeigten moderate Werte für die Mediane; in der Pilotstudie lag der Median bei $m=0,471$, in der abschließenden Analyse bei $m=0,305$. Es ist möglich, dass mit einer dreistufigen Likert-Skala höhere Werte hätten erreicht werden können [28]. Eine deutsche Evaluation der kurzen Version des CCOG zeigte ähnliche Ergebnisse wie die in dieser Publikation diskutierten [16]. Sie berichteten von IRR-Koeffizienten zwischen 0,05 und 0,57.

Eine andere Möglichkeit um die Reliabilität zu verbessern, wäre ein Ratertraining durchzuführen [29]. Obwohl Ratertrainings variieren und in manchen Fällen die Rater nur per E-Mail Informationen erhalten [30], wurde entschieden, kein Ratertraining durchzuführen. Die Entscheidung wurde getroffen, da ein Training, auch wenn es nur in digitaler Art und Weise durchgeführt wird, Zeit benötigt, was den Zielen eines einfachen und zeitlich effizienten Bewertungsbogen widerspricht. Zusätzlich findet häufig kein Ratertraining statt, wenn Fragebögen wie dieser außerhalb des Rahmens einer Studie genutzt werden. Deshalb sollte die vorausgehende Studie ebenfalls ohne ein Ratertraining durchgeführt werden.

Bezüglich Item 13 wurden 45,7% der Videos von beiden Ratern gleich bewertet, während 42,9% der Videos mit einem Punkt unterschied bewertet wurden. Daraus ergibt sich, dass 88,6% der Videos mit nicht mehr als einem Punkt Unterschied bewertet wurden. Ob dies akzeptabel ist, hängt vom Einsatzzweck ab. Für formative Prüfungen und Feedback erscheint es geeignet. Vor dem Einsatz in einer summativen Prüfung sollte man berücksichtigen, ob die Bewertung alleine oder kombiniert mit Bewertungen anderer Bewertungsinstrumente genutzt werden soll. Um komplexe Fertigkeiten wie Kommunikation zu bewerten, sind Globalbewertungen besser geeignet als Checklisten [30], [31]. In dieser Evaluation konnte gezeigt werden, dass die Korrelation zwischen den Checklisten-Items (Items 1 bis 11) und der Globalbewertung (Item 13) sehr hoch ist ($r_1=0,856$, $r_2=0,898$). Zusätzlich wurden

exzellente Werte für die Interne Konsistenz erreicht ($\alpha_1=0,937$, $\alpha_2=0,926$). Auf Grundlagen der Daten könnte geschlussfolgert werden, dass Item 13 als einziges Item ausreichend wäre, um das Eingehen auf den Patienten zu bewerten. Auf der anderen Seite schulen die Checklisten-Items vermutlich den Rater bezüglich der Bewertungskriterien, bevor die Globalbewertungen gegeben wird. Daher bedarf es weiterer Untersuchungen, ob Item 13 als alleiniges Item für summative Prüfungen geeignet ist. Für formative Prüfungen können die Items genutzt werden, um detailreiches Feedback zu geben.

4.3. Limitationen

Betrachtet man die Einschränkungen, muss man zunächst auf die kleinen Stichproben hinweisen. Schon während der Pilotstudie hätte eine größere Datenmenge geholfen, die Items zu verbessern. Insbesondere wäre es ratsam gewesen eine zweite Pilotstudie durchzuführen um herauszufinden, ob die durchgeführten Veränderungen effektiv waren. In der finalen statistischen Analyse wäre es von Vorteil gewesen die Daten einer größeren Anzahl von Ratern zu integrieren, um bessere Werte für die IRR zu erhalten. Wäre die Anzahl der Videos höher gewesen, wäre die statistische Analyse solider ausgefallen.

Eine weitere Einschränkung ist das Ziel das „Eingehen auf den Patienten“ zu messen. Die Entscheidung dieses Konzept zu beurteilen, entwickelte sich aus praktischen Bedürfnissen, obwohl es als gute wissenschaftliche Praxis gilt, ein Konstrukt auf Grundlage von theoretischen Überlegungen oder Literaturrecherche zu wählen. Darüber hinaus gab es keine konzeptuelle Vorgängerstudie, weshalb das Konstrukt nicht gut definiert war. Daher wurde auch keine klare Differenzierung zwischen Empathie, Beziehungsaufbau und Eingehen auf den Patienten getroffen. Andererseits ist es ein Vorteil ein Instrument zu nutzen, das auf praktischen Bedürfnissen basiert.

In der Studie wurden aus praktischen Gründen nur Videos, in denen Gespräche zwischen Studierenden und SPs aufgezeichnet waren, bewertet. Es könnte sein, dass die Beobachtung eines direkten Kontaktes zwischen echten Patient*innen und Studierenden andere Ergebnisse zeigen würde. Es wird davon ausgegangen, dass die Unterschiede zwischen SPs und richtigen Patient*innen klein sind, wenn die SPs gut schauspielern. Der Unterschied zwischen auf Video aufgenommenen Begegnungen und direkten Begegnungen ist schwierig abzuschätzen, da es zu Unterschieden kommen kann, in dem was Rater beobachten können, wie z.B. die Gestik, die Mimik wie auch die Atmosphäre zwischen den beiden Parteien. In

dieser Studie war es den Ratern nicht gestattet Szenen oder Videos erneut anzusehen, da dies vermutlich zu einem Unterschied im Vergleich zu einer direkten Bewertung geführt hätte. Zusätzlich wäre bei einer Verwendung außerhalb des Rahmens einer Studie vermutlich nicht genügend Zeit, die Videos oder auch nur Teile davon, mehrfach anzusehen.

Das Ziel war, einen Bewertungsbogen zu entwickeln, der einfach zu nutzen ist. Dennoch ist es notwendig sich einige Minuten Zeit zu nehmen um mit den Items vertraut zu werden bevor das Gespräch beobachtet wird. Da es fundamental ist, das gesamte Gespräch zu beobachten, bevor die Bewertung vorgenommen wird, wird im Anschluss erneut Zeit benötigt, um die Bewertung durchzuführen. Um das Instrument möglichst zeiteffizient zu gestalten, wurden die Daten, die aus dieser Studie gewonnen wurden, ohne Ratertraining generiert.

Eine weitere Einschränkung ergibt sich bezüglich der Sprache. In dieser Studie wurde die Validierung anhand einer deutschen Version durchgeführt. Die Übersetzung ins Englische erfolgte nur für die Veröffentlichung. Zusätzlich wird empfohlen verschiedene Schritte während des Übersetzungsprozess, wie z. B. eine Rückübersetzung und diverse Überprüfungen, durchzuführen [32]. In diesem Fall wurde der Bewertungsbogen von der Autorin und einem Englischmuttersprachler, beide mit exzellenten Kenntnissen der anderen Sprache, ohne weiteren Übersetzungsprozess übersetzt.

4.4. Ausblick

Zukünftige Studien könnten analysieren, ob die Bewertung des „Eingehens auf den Patienten“ durch Patient*innen oder SPs mit der Bewertung einer nicht beteiligten beobachtenden Person korreliert. Hierfür kann Item 12 (Globalbewertungen aus Patienten Perspektive) genutzt werden.

Dieser Bewertungsbogen kann zukünftig in einem breiten Feld eingesetzt werden. Wie oben beschrieben, kann das Instrument in der Lehre eingesetzt werden sowie in Prüfungen, wobei dabei bedacht werden sollte, dass die Verwendung in summativen Prüfungen kritisch geprüft werden muss. In dieser Studie wurde er nur genutzt, um auf Video aufgenommene Situationen zu bewerten, es ist jedoch auch vorstellbar, dass er ebenfalls in direkten Beobachtungen eingesetzt wird. Es wird angenommen, dass er in allen Stufen der medizinischen Ausbildung eingesetzt werden kann. Er wurde in zwei Situationen während des Studiums [Vorklinik und Klinik] (Pilotstudie und abschließende statistische Analyse) überprüft, aber ein Einsatz im Bereich der Weiterbildung ist ebenfalls vorstellbar.

Wahrscheinlich werden die Meisten zunächst an einen Einsatz des Bewertungsbogens in einem Fach, in dem der Fokus Kommunikation mit Patient*innen liegt, denken. Er kann jedoch auch in anderen Fächern angewandt werden. Die Themen der Gespräche, in den aufgenommenen Videos der finalen statistischen Analyse, standen im Zusammenhang mit der Lehre in der Chirurgie. Obwohl

der Bogen in allen Situationen angewandt werden kann, ist vermutlich am hilfreichsten, ihn im Training oder der Prüfung von fordernden Gesprächssituationen einzusetzen.

5. Zusammenfassung

Die psychometrischen Eigenschaften des Bewertungsbogens „Eingehen auf den Patienten – Marburger Bewertungsbogen“ legen nahe, dass er gut für die Lehre und formative Prüfungen genutzt werden kann. Seine Anwendung in summativen Prüfungen kann in Erwägung gezogen werden. Nichtsdestotrotz sollte evaluiert werden, ob eine dreistufige Likert-Skala höhere Werte der IRR erreichen könnte. Darüber hinaus wäre es interessant zu untersuchen, ob Item 13 als eigenständiges Item genutzt werden könnte, da dies kontrovers diskutiert wird.

Danksagung

Wir danken allen Studierenden und Ratern, die zum Erfolg der Studie beigetragen haben. Besonders wollen wir Lorianna Könniker für ihre Hilfe bei der Übersetzung des Bewertungsbogens danken. Außerdem wollen wir die finanzielle Unterstützung der Abteilung für Allgemeinmedizin der Philipps-Universität anerkennen, die es möglich machte, dass die Rater eine Aufwandsentschädigung erhalten konnten; dem Förderverein Maris e.V., der die Kosten des Proofreading Services übernahm und dem Skill-Lab „Maris“ selbst für die Übernahme der Löhne der Tutoren danken wir.

Steckbrief

Name des Standorts: Philipps-Universität Marburg

Studienfach/Berufsgruppe: Medizin

Anzahl der Lernenden pro Jahr: 350 p.a.

Ist ein longitudinales Kommunikationscurriculum implementiert? Ja

In welchen Semestern werden kommunikative und soziale Kompetenzen unterrichtet? 1, 2, 5, 6, 7, 8, 9, (PJ)

Welche Unterrichtsformate kommen zum Einsatz? Seminare, praktische Übungen, SP-Einsatz/Simulation

In welchen Semestern werden kommunikative und soziale Kompetenzen geprüft? Formativ: 2 und 9, bestehensrelevant und benotet 5 und 6

Welche Prüfungsformate kommen zum Einsatz? Rollenspiel+Videofeedback, OSCEs mit SP

Wer (z.B. Klinik, Institution) ist mit der Entwicklung und Umsetzung betraut? Interdisziplinäres Skills Lab in Kooperation mit Fächern/Instituten

Aktuelle berufliche Rolle der Autor*innen

- Dr. med. Miriam Urff studierte Medizin in Marburg und ist Assistenzärztin am Universitätsklinikum Freiburg, Klinik für Psychiatrie, Psychotherapie und Psychosomatik im Kindes- und Jugendalter. Sie war drei Jahre lang studentische Tutorin im Marburger interdisziplinären Skills Lab und promovierte mit einem Projekt zur ärztlichen Gesprächsführung in der Chirurgie.
- Prof. Dr. med. Antonio Krüger ist Chefarzt der Klinik für Unfallchirurgie, Orthopädie, Wirbelsäulen- und Kindertraumatologie an der Asklepios Klinik in Lich. Er war langjähriger Oberarzt an der Universitätsklinik Marburg, PJ- und Lehrbeauftragter sowie Vertrauensdozent. Er ist überzeugt davon, dass die Lehre interdisziplinär – zum Beispiel durch Lernzentren – verbessert werden kann.
- Prof. Dr. med. Steffen Ruchholtz ist seit 2015 geschäftsführender Direktor am Zentrum für Orthopädie und Unfallchirurgie am Universitätsklinikum Marburg. Als Sprecher des Ausschusses TraumaNetzwerk der Deutschen Gesellschaft für Unfallchirurgie ist er außerdem federführend für die Umsetzung dieses Konzeptes zur Optimierung der klinischen Behandlungsqualität schwerverletzter Patienten in Deutschland verantwortlich.
- Dr. phil. Tina Stibane leitet das Zentrum für medizinische Lehre an der Philipps-Universität Marburg. Sie koordiniert das Skills Lab Maris, betreut verschiedene Lehr-Lernprojekte und ist konzeptionell und moderierend in der Curriculumskommission tätig. Sie führt medizin-didaktische Schulungen für Dozent*innen und Tutor*innen am Fachbereich Medizin durch.

Daten

Daten für diesen Artikel sind im Dryad-Repository verfügbar unter: <http://dx.doi.org/10.5061/dryad.qjq2bvqcw> [33]

Interessenkonflikt

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

Literatur

1. Kurtz S, Silverman J, Draper J. Teaching and Learning Communication Skills in Medicine. 2nd ed. Oxford, San Francisco: Radcliffe Publishing; 2005.
2. Schweickhardt A, Fritzsche K. Kursbuch ärztliche Kommunikation: Grundlagen und Fallbeispiele aus Klinik und Praxis; mit 15 Tabellen. 2nd ed. Köln: Deutscher Ärzte-Verlag; 2009.
3. Berian JR, Ko CY, Angelos P. Surgical Professionalism: The Inspiring Surgeon of the Modern Era. *Ann Surg.* 2016;263(3):428-429. DOI: 10.1097/SLA.0000000000001425
4. Smith RC, Lyles JS, Mettler J, Stoffelmayr BE, Van Egeren LF, Marshall AA, Gardiner JC, Maduschke KM, Stanley JM, Osborn GG, Shebroe V, Greenbaum RB. The effectiveness of intensive training for residents in interviewing. A randomized, controlled study. *Ann Intern Med.* 1998;128(2):118-126. DOI: 10.7326/0003-4819-128-2-199801150-00008
5. Levinson W, Roter D. The effects of two continuing medical education programs on communication skills of practicing primary care physicians. *J Gen Intern Med.* 1993;8(6):318-324. DOI: 10.1007/BF02600146
6. Yedidia MJ, Gillespie CC, Kachur E, Schwartz MD, Ockene J, Chepaitis AE, Snyder CW, Lazare A, Lipkin Jr M. Effect of communications training on medical student performance. *JAMA.* 2003;290(9):1157-1165. DOI: 10.1001/jama.290.9.1157
7. Kiessling C, Dieterich A, Fabry G, Hölzer H, Langewitz W, Mühlhans I, Pruskil S, Scheffer S, Schubert S. Basler Consensus Statement "Kommunikative und soziale Kompetenzen im Medizinstudium": Ein Positionspapier des GMA-Ausschusses Kommunikative und soziale Kompetenzen. *GMS Z Med Ausbild.* 2008 Mai;25(2):Doc83. Zugänglich unter/available from: <https://www.egms.de/static/de/journals/zma/2008-25/zma000567.shtml>
8. Frank JR, editor. The CanMEDS 2005 physician competency framework Better standards. Better physicians. Better care. Ottawa: The Royal College of Physicians and Surgeons of Canada; 2005.
9. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. *Acad Med.* 2001;76(4):390-393. DOI: 10.1097/00001888-200104000-00021
10. Schirmer JM, Mauksch L, Lang F, Marvel MK, Zoppi K, Epstein RM, Brock D, Pryzbylski M. Assessing communication competence: a review of current tools. *Fam Med.* 2005;37(3):184-192.
11. Sennekamp M, Gilbert K, Gerlach FM, Guethlin C. Development and validation of the "FROCK": Frankfurt observer communication checklist. *Z Evid Fortbild Qual Gesundhwes.* 2012;106(8):595-601. DOI: 10.1016/j.zefq.2012.07.018
12. Simmenroth-Nayda A, Weiss C, Fischer T, Himmel W. Do communication training programs improve students' communication skills? - a follow-up study. *BMC Res Notes.* 2012;5:486. DOI: 10.1186/1756-0500-5-486
13. Norfolk T, Birdi K, Walsh D. The role of empathy in establishing rapport in the consultation: a new model. *Med Educ.* 2007;41(7):690-697. DOI: 10.1111/j.1365-2923.2007.02789.x
14. Hojat M. Empathy in Health Professions Education and Patient Care. 1st ed. Berlin, Heidelberg: Springer; 2016. DOI: 10.1007/978-3-319-27625-0
15. Silverman J, Kurtz S, Draper J. Skills for Communicating With Patients. 3rd ed. London New York: Radcliffe Publishing; 2013.
16. Simmenroth-Nayda A, Heinemann S, Nolte C, Fischer T, Himmel W. Psychometric properties of the Calgary Cambridge guides to assess communication skills of under-graduate medical students. *Int J Med Educ.* 2014;5:212-218. DOI: 10.5116/ijme.5454.c665
17. Bosse HM, Schultz JH, Nickel M, Lutz T, Möltner A, Jünger J, Huwendiek S, Nikendei C. The effect of using standardized patients or peer role play on ratings of undergraduate communication training: a randomized controlled trial. *Patient Educ Couns.* 2012;87(3):300-306. DOI: 10.1016/j.pec.2011.10.007
18. Choudhary A, Gupta V. Teaching communications skills to medical students: Introducing the fine art of medical practice. *Int J Appl Basic Med Res.* 2015;5(Suppl 1):S41-44. DOI: 10.4103/2229-516X.162273

19. Nguyen LN, Tardioli K, Roberts M, Watterson J. Development and incorporation of hybrid simulation OSCE into in-training examinations to assess multiple CanMEDS competencies in urologic trainees. *Can Urol Assoc J.* 2015;9(1-2):32-36. DOI: 10.5489/cuaj.2366
20. Roh H, Park KH, Jeon YJ, Park SG, Lee J. Medical students' agenda-setting abilities during medical interviews. *Korean J Med Educ.* 2015;27(2):77-86. DOI: 10.3946/kjme.2015.27.2.77
21. Ammentorp J, Thomsen JL, Jarbøl DE, Holst R, Øvrehus AL, Kofoed PE. Comparison of the medical students' perceived self-efficacy and the evaluation of the observers and patients. *BMC Med Educ.* 2013;13:49. DOI: 10.1186/1472-6920-13-49
22. Harasym PH, Woloschuk W, Cuning L. Undesired variance due to examiner stringency/leniency effect in communication skill scores assessed in OSCEs. *Adv Health Sci Educ Theory Pract.* 2008;13(5):617-632. DOI: 10.1007/s10459-007-9068-0
23. Schnell R, Hill PB, Elke E. *Methoden der empirischen Sozialforschung.* 10th ed. München, Oldenbourg: De Gruyter Oldenbourg; 2013.
24. Wirtz M, Caspar F. *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen.* Göttingen, Bern, Toronto, Seattle: Hogrefe - Verlag für Psychologie; 2002.
25. Kühnel SM, Krebs D. *Statistik für die Sozialwissenschaften: Grundlagen, Methoden, Anwendungen.* 6th ed. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag; 2012.
26. Schmitt N. Uses and Abuses of Coefficient Alpha. *Psychol Ass.* 1996;8(4):350-353. DOI: 10.1037/1040-3590.8.4.350
27. Kline P. *The Handbook of Psychological Testing.* London, New York: Routledge; 1993.
28. Nolte C. *Validität und Reliabilität eines Instruments zur Messung der Qualität der Kommunikation und seine Eignung im studentischen Unterricht.* Göttingen: Georg-August-Universität zu Göttingen; 2014.
29. Crossley J, Humphris G, Jolly B. Assessing health professionals. *Med Educ.* 2002;36(9):800-804. DOI: 10.1046/j.1365-2923.2002.01294.x
30. Scheffer S. *Validierung des "Berliner Global Rating" (BGR) - ein Instrument zur Prüfung kommunikativer Kompetenzen Medizinstudierender im Rahmen klinischpraktischer Prüfungen (OSCE).* Berlin: Charité - Universität Berlin; 2009.
31. Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ.* 2004;38(2):199-203. DOI: 10.1111/j.1365-2923.2004.01755.x
32. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P; ISPOR Task Force for Translation and Cultural Adaptation. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health.* 2005;8(2):94-104. DOI: 10.1111/j.1524-4733.2005.04054.x
33. Urff M, Krüger A, Ruchholtz S, Stibane T. "Catering to the patient" - psychometric properties of an innovative assessment instrument. *Dryad Digital Repository.* 2021. DOI: 10.5061/dryad.qjq2bvqcw

Korrespondenzadresse:

Dr. Eva Christina Stibane
 Philipps-Universität Marburg, Zentrum für Medizinische
 Lehre, Conradstr. 7, 35043 Marburg, Deutschland, Tel.:
 +49 (0)6421/28-25053
 stibane@staff.uni-marburg.de

Bitte zitieren als

Urff M, Krüger A, Ruchholtz S, Stibane EC. *Catering to the patient – development, validation and psychometric properties of an innovative assessment instrument.* *GMS J Med Educ.* 2021;38(3):Doc69. DOI: 10.3205/zma001465, URN: urn:nbn:de:0183-zma0014650

Artikel online frei zugänglich unter

<https://www.egms.de/en/journals/zma/2021-38/zma001465.shtml>

Eingereicht: 31.03.2020
Überarbeitet: 14.07.2020
Angenommen: 21.09.2020
Veröffentlicht: 15.03.2021

Copyright

©2021 Urff 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/>.