

# Development and testing of a summative video-based e-examination in relation to an OSCE for measuring communication-related factual and procedural knowledge of medical students

## Abstract

**Objective:** In the context of educating medical students, testing of competence in medical communication is carried out primarily with the Objective Structured Clinical Examination [1]. This makes it possible to assess practical performance, but it is resource-intensive and has a negative impact on test quality. The project “Digital test tool for measuring communication skills in medical studies” (digjRole) was funded by the BMBF (Federal Ministry of Education and Research) and its objective was to develop digital formats as electronic versions of an OSCE in order to test the communication competency of medical students. Such digital forms of examination should be cost-effective, be relevant to clinical practice and have high psychometric quality. In terms of content, the examination questions should incorporate factual and procedural knowledge as components of communication competency, although we assumed that procedural knowledge is more relevant than factual knowledge to OSCE performance. This article describes the development and testing of a video-based, communication-related e-examination that is relevant to passing the test, which is the first milestone of the overall project.

**Methodology:** We produced videos and related exam questions in the form of a situational judgement test [2] related to medical psychology and medical sociology, based on the educational content of a preclinical course on doctor-patient communication at the Mainz University Medical Center. In the summer semester of 2018, 226 students sat for this video-based single-choice e-examination (VSE). In the winter semester of 2018/2019, a different cohort of 192 students participated in the VSE as well as a tried-and-tested communication OSCE with five stations [3].

**Results:** The internal consistencies for the VSE in the summer semester of 2018 were  $\alpha=.55$ , in the winter semester 2018/19 with  $\alpha=.62$  and for the OSCE with  $\alpha=.60$ . There was a positive correlation between the performance of the students with the VSE and that with the OSCE ( $r=.21$ ,  $p\leq.01$ ). Principal Axis Analyses did not reveal any dimensioning in terms of factual and procedural knowledge. In the evaluation, the majority of the students stated that the VSE was quite relevant to the practise of medical communication and were in favour of retaining this form of examination.

**Conclusion:** The correlation between the VSE and the OSCE is relatively low, so that the VSE in this form is not a satisfactory predictor of an OSCE result. In terms of internal consistency, the VSE and the OSCE produced an almost identical result. It can also be assumed that the VSE can achieve a high degree of objectivity with the use of standardised video-based examinations as well as greater resource efficiency than OSCEs.

**Keywords:** medical education, communicative competencies, summative assessment, video-based examination, OSCE

Stephanie Ludwig<sup>1</sup>

Lina Behling<sup>2</sup>

Uwe Schmidt<sup>2</sup>

Sabine Fischbeck<sup>1</sup>

1 University Medical Center,  
Johannes Gutenberg  
University Mainz, Clinic and  
Polyclinic for Psychosomatic  
Medicine and Psychotherapy,  
Department of Medical  
Psychology and Medical  
Sociology, Mainz, Germany

2 Johannes Gutenberg  
University Mainz, Center for  
Quality Assurance and  
Development (ZQ), Mainz,  
Germany

# 1. Introduction

## 1.1. State of research

Teaching and testing communication skills is becoming increasingly relevant in the context of studying medicine. The “Master Plan for Medical Education 2020” emphasises the importance of relevant training as part of the medical course [4]. Education in communication skills is recognised internationally as an essential part of medical training [5], [6]. With regard to styles of teaching and examination relevant to competence in medical communication, Härtl et al. [1] surveyed universities in German-speaking countries and found that teachers use models or catalogues of learning objectives in 70% of degree programmes. The concept of competence in medical communication is regarded as complex and multi-layered [7], [8]. The review article [1] for the German-speaking area and one by Laidlaw et al. [9] for Great Britain also show that, with regard to the types of examination, testing competence in communication takes place primarily in the form of the practical OSCE (Objective Structured Clinical Examination). The written test is the second most common form in German-speaking countries. In other countries such as Australia, Canada and the USA, the OSCE style of examination is also used in addition to classic written examinations [10], [11], [12]. With the OSCE, the student’s competence is assessed by the examiners, who mostly use checklists based on simulated situations at several examination stations - sometimes with the aid of simulated patients. When comparing written forms of examination, the advantage of an OSCE is its high practical relevance and the possibility of a performance test at a behavioural level [13]. By contrast, written exams are at the cognitive level, and can be broken down into factual knowledge and procedural knowledge [13]. This level can be checked, for example, with the aid of single or multiple choice questions, key feature concepts (making critical decisions) or situational judgement tests [2], [14], [15], [16]. In a situational judgement test, scenarios from professional practice are presented in writing or video-based. An appropriate action must be decided in the context of the examination questions. The response can be formulated in various ways. The internal consistencies are mostly good [2]. Situational judgement tests are already used in medical training [15]. In terms of economy, implementation of an OSCE is very time consuming and resource-intensive with regard to financial and personnel aspects. The reliability and content validity of an OSCE is largely determined by the number of stations [17], which makes the effort enormous and often unaffordable in the desired form. Assessment by various auditors may provide different results, so that inter-rater reliability (and therefore objectivity) may be inadequate [17]. However, a written test (such as a situational judgement test) generally achieves greater objectivity and reliability due to the standardisation of questions and evaluation as well as a greater number of questions.

The presented advantages and disadvantages of the two most common ways of examining competence in medical communication raise the question of alternative forms of examination that take into account the importance of a high practical relevance as well as the requirements of a realistic resource procedure with a high test quality. Digital formats with videos appear to be ideal in the field of communication competence, since they enable case-based testing with a high degree of practical relevance and a large number of students. This is consistent with student feedback from the classroom that doctor-patient consultations shown in videos are much closer to a real-life situation than just written case descriptions, in which images, sound, facial expressions and gestures are completely missing. This makes it easier for students to imagine themselves in the role of a doctor and “experience” the situation than it would be with just written exams. There is also the option of incorporating questions and videos into exam software, so that the examination can be designed economically and objectively. Application of digital formats to teaching competence in medical communication is familiar to the authors, mainly in relation to video-based exercises for learning about competence in communication when this project started in German-speaking areas in 2017 [18]. With regard to examinations, the only report of the use of videos in examinations (at two universities) is that of Härtl et al. [1]. Several relevant studies have been carried internationally that deal with the development and testing of video-based examinations for measuring skills in medical communication [8], [19], [20]. At universities where these studies were carried out, an Objective Structured Video Exam (OSVE) was used, in which students were given tasks related to videos on doctor-patient consultations, which included questions in short essay format and naming communication skills. Humphris and Kaney [19] found a correlation of  $r=.17$  between the OSVE and the OSCE.

## 1.2. Project design

The lack of resource-saving, competence-oriented exams with significant practical relevance led us to the idea of developing, testing and scientifically investigating a video-based examination of factual and procedural knowledge as components of communication skills in medical studies for the first time in German-speaking regions. The exam should be regarded as a digital variant of the OSCE. In the first phase of the study, the video-based examination was designed as a summative examination (video-based single-choice examination, VSE) with tasks in the form of a situational judgement test. For the reasons outlined above, we assume that the VSE can achieve higher practical relevance than purely written examinations (also felt by the students), proves to be economical, and achieves a high level of implementation and objectivity of evaluation (with software management). The high level of economy is expected because, although much time and money need to be invested in creating the scenarios and producing the videos, there should be enough mater-

ial to enable students to be tested inexpensively over several semesters. This makes it possible to reduce the amount of time taken by the examiners with the OSCE. It is also assumed that there is a positive correlation between the examination performance of the students with the VSE and their performance with the OSCE (cf. also [19] and [21] with  $r=.32$  for the correlation of key feature testing and the OSCE). The underlying dimensions of the VSEs will be examined, and it is conceivable that the design of the examination is reflected in a two-factor solution across all content-related topics in accordance with the two aspects “factual knowledge” and “procedural knowledge”. Should such a solution arise, one should also find out whether the questions in the VSE at the level of procedural knowledge are more closely related to the results of the OSCE than the questions on the level of factual knowledge, since it is obvious that procedural knowledge and the action taken seems to be more closely related than factual knowledge and the action taken. All examinations should be evaluated by the students in order to assess their practical relevance and whether or not this form of examination should be maintained.

## 2. Methods

### 2.1. Study participants and process

A VSE was used for the first time in the course II "Doctor-Patient Communication" for students studying medicine in the second semester at the Mainz University Medical Center in the summer semester of 2018 (SS18). In the winter semester of 2018/19 (WS18/19), the students of the following cohort in the aforementioned course in the second semester also completed the VSE and then the OSCE. Prior to implementation, the courses included explanations of the precise design of the exams with a sample task. There was also a detailed information sheet that could be read separately. The VSEs were carried out on the premises of the Centre for Data Processing at the University of Mainz at the end of the semester. The technical implementation took place with ILIAS testing software. Except for a few problems with the sound quality and the loading times of the videos, which were easily resolved, implementation of the VSEs ran smoothly without any problems. After the examination, the evaluation was carried out with paper-and-pencil surveys. The higher response rate in the evaluation of the VSE in the second cohort is probably due to the fact that the examination supervisors referred more to evaluation based on experiences from the previous semester and collected it in a targeted manner. Information on the number, age and gender of the study participants is given in table 1.

### 2.2. Development of the VSE

In the development phase of the VSE, which was based on the examination content from previous years, case reports and dialogues between doctor and patient were

developed in an initial step. The dialogues contain the essential theoretical concepts that students should be able to apply in practice (including the Calgary-Cambridge scheme [22], NURSE model [23], counselling techniques such as active listening [24], principles of participatory decision-making [25] and the SPIKES protocol [26]). The scenarios incorporate the following requirements: checking and promoting drug compliance, discussing medical histories, communicating a cancer diagnosis according to the SPIKES protocol, conducting ward rounds, discussing therapy for hypertension according to the principle of participatory decision-making with the patient, changing the behaviour of obese patients, analysing stress reactions and conveying information. The dialogues were checked and revised by course instructors. Interviews with experts (including general practitioners and psychologists) were also carried out to ensure authenticity. Experts were selected on the basis that they had practical experience in everyday working life, were perhaps involved in teaching themselves or, in one case, had been involved in the development of a previous OSCE. The experts gave written feedback on the dialogues (based on the questions asked, e.g. requesting an assessment of the credibility of the setting, medical history, conducting a consultation, presenting a patient etc.). These were evaluated qualitatively and the dialogues were revised accordingly.

The videos were produced in collaboration with Mainz University's media centre and the learning clinic of Mainz University's Medical Center. Simulated patients and a general practitioner took the appropriate roles. A split-screen process was used to process the video content (both long and frontal shots) so that assistance with changing perspectives could be avoided. Consistent with the recommendation of Hoppe-Seyler et al. [27], the students opinion on the credibility of the shown scenes was regarded as important, therefore a trial production of a scene was made prior to the final filming. This was evaluated by eight students (selected from an e-mail distribution list for medical students from different semesters). The “doctor” role (63%), “patient” role (75%), “practice” (88%) and “interaction” (88%) were assessed by the majority of the students as realistic.

The questions related to the exam videos were video-adapted and developed in a single-choice format, each with five answer options and reference to existing exam questions. The tasks require either factual knowledge of the presented sequence or a practical decision/choice of a suitable formulation in the sense of a situational judgement test. This procedure takes into account the course design with theoretical and practical content. In addition, it was decided to question the wide range of topics covered in the course and not to narrow it down to specific topics. In the first exam, the focus was slightly more factual knowledge and in the second exam slightly more on procedural knowledge (for the breakdown into factual and procedural knowledge, see table 2). The questions were checked and revised by course instructors.

**Table 1: Study participants**

	VSE SS18	VSE WS18/19	OSCE WS18/19	OSCE and VSE WS18/19
<b>N</b>	226	193	195	192
<b>Age</b>	M = 23 SD = 4.3 (18–45)	M = 25 SD = 3.8 (19–49)*	M = 25 SD = 3.8 (19–49)*	M = 25 SD = 3.8 (19–49)*
<b>Female</b>	58%	58%	57%	58%
<b>Male</b>	42%	42%	43%	42%

Comment. \* 2 missing.

**Table 2: Item characteristics of the VSEs in SS18 (N=226) and in WS18/19 (N=193)**

VSE SS18				VSE WS18/19			
	correct	wrong	Item-total correlation $r_{it}$		correct	wrong	Item-total correlation $r_{it}$
<b>Item (main topics)</b>				<b>Item (main topics)</b>			
<b>Items – procedural knowledge</b>							
4 Patient orientation / empathy (with non-compliance) *	202	24	.30	22 Patient orientation/empathy (with non-compliance) *	174	19	.06
11 Patient orientation / empathy *	220	6	.33	23 Patient orientation/ empathy *	189	4	.17
16 Patient orientation / empathy *	213	13	.31	25 Patient orientation/ empathy *	190	3	.30
19 Patient orientation / empathy (during ward rounds) *	214	12	.27	26 Patient orientation/ empathy (during ward rounds) *	181	12	.26
1 Comprehensible communication of information	212	14	.16	2 Risk communication	89	104	.15
7 Stress analysis	109	117	.14	3 Patient orientation/empathy (positive appreciation)	184	9	.36
9 SPIKES protocol	208	18	.10	5 Patient orientation/empathy	175	18	.16
12 Participatory decision-making	159	67	.18	7 Patient orientation/empathy (motivational interviewing)	155	38	.29
15 Behaviour modification	139	87	.08	8 Comprehensible communication of information	169	24	.07
20 Ward rounds, structure, information	224	2	.14	9 Patient orientation/empathy (asymmetrical avoidance behaviour)	147	46	.32
25 Managing side effects	135	91	.08	10 Participatory decision making	32	161	.19
27 Facilitators	225	1	-.04	14 Medical history	155	38	.18
30 Comprehensible communication of information	57	169	-.01	15 Medical history	166	27	.22
				20 Compliance	179	14	.24
				21 Motivational interviewing	182	11	.30
				29 Active listening	122	71	.24
				30 Patient orientation / empathy ('difficult' patient)	182	11	.44
<b>Items - factual knowledge</b>							
13 Patient orientation/ empathy (Schulz von Thun) *	219	7	.17	24 Patient orientation/ empathy (Schulz von Thun) *	186	7	.26
2 Medical history	187	39	.15	1 SPIKES protocol	86	107	.08
3 Question types	169	57	.20	4 Operant conditioning	145	48	.26
5 Medical history	150	76	.14	6 "Stages of change" model	150	43	.04
6 Metacommunication	100	126	.01	11 Patient role in the general hospital	147	46	.26
8 Active listening	204	22	-.01	12 Patient orientation/empathy	179	14	.34
10 Nonverbal communication	185	41	.26	13 Doctor-patient relationship	180	13	.24
14 "Stages of change" model	160	66	.18	16 Assessment errors	185	8	.23
17 Assessment errors	220	6	.04	17 NURSE model	189	4	.12
18 Participatory decision-making	187	39	.19	18 Question types	161	32	.14
21 NURSE model	210	16	.09	19 Non-verbal communication	157	36	.11
22 Asymmetric avoidance behaviour	195	31	.21	27 Behaviour modification	119	74	.05
23 Empathy theory	20	206	.11	28 Concept of control belief	101	92	.04
24 Transactional stress model	126	100	.13				
26 Patient orientation/ behaviour modification	193	33	.26				
28 WHO ICF model	113	113	.21				
29 Motivational interviewing	115	111	.12				

Comment. The items marked with \* were used in both exams. With regard to the other items, they are partly the same main topic across all exams, but the contents of the task and of the questions are different. 'Correct' and 'wrong' are frequency data.



The examination tasks were held in the following style. The video shows the case of a 50-year-old patient who was diagnosed with hypertension and for whom a treatment decision must be made according to the principle of participatory decision-making. The video shows how the female doctor tells the patient that it is important to her to make a decision about the treatment steps together with him, and the patient agrees. The subsequent task reads "The doctor and patient have now agreed on equal rights with regard to the decision. When implementing participatory decision-making in doctor-patient discussions, a sequence of steps must be taken into account. Which step follows the model of participatory decision-making?" The five possible answers are "Say that a decision is pending", "Inform about options" (correct answer), "Provide information about the advantages and disadvantages of the options", "Inquire about understanding, thoughts and expectations" and "Determine preferences".

The evaluation was carried out by dichotomising the five possible answers (correct/incorrect) and then determining a total value. According to the results of the Principal Axis Analysis reported below, five items from SS18 were used again in WS18/19.

### 2.3. OSCE

The OSCE was checked on the basis of many years of experience with a communication OSCE in Mainz [3] derived from five stations that have already been developed (stations on the main topics of medical history, analysis of the stress reaction, checking and promoting compliance, participatory decision-making and notification of a cancer diagnosis according to the SPIKES protocol). The performance of the students was assessed by a total of 11 examiners over three examination days, using matching checklists that have already been developed and tested. The checklists related to the specific contents of the stations as well as aspects of the counselling techniques such as empathy, active listening, eye contact and open body position (for example "took an open body position"). The contents therefore correspond in the essential topics with the contents of the VSEs, although further topics and more theoretical knowledge were added. Performance was evaluated with a pre-determined weighting of the aspects of the checklists, with a maximum of 7 points for each station. An overall maximum of 35 points could therefore be achieved (absolute limit for a pass 60%).

### 2.4. Statistical evaluation

The internal reliability of each examination was determined using Cronbach's Alpha. Spearman's rank correlation coefficient was used to examine the relationship between students' performance in the VSE and OSCE examinations. Kolmogorov-Smirnov tests were used to test the normal distribution of the variables. An iterative Principal Axis Analysis was performed on the VSE questions from

each of the academic terms. Since the items under consideration are dichotomous variables that also show partly very low and very high item difficulties (see results section and table 2) and unequal marginal sum distributions are therefore present, odds ratios were calculated between all items. Odds ratios have values between zero and infinity and were transformed to values between -1 to +1 using the formula proposed by Yules. The iterative Principal Axis Analysis is based on the transformed odds ratios matrix. The Yules Y formula was applied with  $Y = (\sqrt{OR} - 1) / (\sqrt{OR} + 1)$  [28], [29], [30].

## 3. Results

The item characteristics for both VSEs are shown in table 2. The majority of items were answered correctly by most students ( $M = .75$  SS 18 and  $M = .80$  WS 18/19 across all means). The item-total correlation of the VSEs items ranged from -.04 to .44.

The criteria of Rost and Schermer were used to extract the number of relevant factors [31]. In the SS 18 group, 13 eigenvalues were greater than 1, whereas only one factor appeared to be plausible based on the scree plot. Only variables that had a communality of  $h^2 \geq .16$  and for which the absolute value was  $a \geq .40$  were represented by this factor (see table 3). The remaining 8 items are difficult to represent accurately with a single factor. However, prioritising content plausibility over mathematical solution, merging 5 items (item 4, 11, 13, 16, 19) results in a significantly more appropriate content alignment under consideration of "patient-oriented, empathetic action and understanding based on the patient's situation and the content of the conversation" (Cronbach's  $\alpha = .62$ ,  $N = 226$ ). In the WS 18/19 group, 12 eigenvalues greater than 1 also resulted in only one factor based on the scree plot, whereas 14 items remained after applying the criteria. Here, 10 of the 14 items were also assigned to the aforementioned topic (3, 5, 7, 9, 12, 23, 24, 25, 26, 30; Cronbach's  $\alpha = .60$ ,  $N = 193$ ). Four further items show overlaps in content with the topic, but also with other topics, so that they could not be clearly assigned to the factor. Overall, it should be noted that the allocations and separations are not fully and finally accurate when considering the content of all items. Of the 5 items from the SS 18 group, 4 items were loaded again on the factor in the WS 18/19 group (items 23-26 WS 18/19).

The internal consistency of the VSE in the SS 18 group resulted in a Cronbach's Alpha coefficient of  $\alpha = .55$  ( $N = 226$ ) and in the WS 18/19 group of  $\alpha = .62$  ( $N = 193$ ). For the OSCE, there was an internal consistency of  $\alpha = .60$  ( $N = 195$ ).

The median of the VSE in SS18 was  $Mdn = 23$ ,  $IQR = 3$  for  $N = 226$ , for the VSE in WS18/19 there was a  $Mdn = 25$ ,  $IQR = 3$  for  $N = 192$  and for the OSCE, a  $Mdn = 26.25$ ,  $IQR = 2.75$  also for  $N = 192$ . For the WS18/19 group, the Spearman's rank correlation coefficient with respect to VSE and OSCE was  $r = .21$  ( $p \leq .01$ ). The two variables were not normally distributed (Kolmogorov-Smirnov test OSCE

**Table 3: Principal axis analyses of the VSEs in SS18 and WS18/19, single factor solutions**

	VSE SS18		VSE WS18 / 19	
	Load a	Communality h <sup>2</sup>	Load a	Communality h <sup>2</sup>
	Factor: Patient orientation/empathy		Factor: Patient orientation/empathy	
Item				
1	.56	.31		
2				
3			.75	.56
4	.64	.40		
5			.42	.18
6				
7			.44	.19
8				
9			.54	.29
10	.45	.21		
11	.85	.72		
12			.65	.42
13	.66	.44	.50	.25
14				
15				
16	.81	.65	.63	.40
17				
18				
19	.66	.44		
20			.52	.27
21			.61	.37
22				
23			.50	.25
24			.58	.34
25			.76	.58
26	.47	.22	.60	.36
27				
28				
29				
30			.77	.59

*Comment.* Loads of items are given for which  $h^2 \geq .16$  and absolute amount  $a \geq .40$ .

$D_{192} = .09$  and VSE WS 18/19  $D_{192} = .15$ , both  $p \leq .01$ ). Furthermore, an additional examination of the correlation of the items exclusively related to procedural knowledge from the VSE and OSCE results yielded a Spearman's rank correlation coefficient of  $r = .25$  ( $p \leq .01$ ). The variable "procedural knowledge" was also not normally distributed (Kolmogorov-Smirnov test:  $D_{192} = .19$ ,  $p \leq .01$ ).

The results of the evaluation of the VSEs are presented in table 4, with a majority of the students in favour of continuing the VSE and rating the practical relevance as relatively high. As the OSCE had already been tested and evaluated, we note that this was also rated as "good" in the WS 18/19 group with an overall score of  $Mdn = 2$ ,  $IQR = 1$  ( $N = 145$ ).

## 4. Discussion

We realised the goal of initial development and testing of a video-based e-exam for testing factual and procedural knowledge as components of competence in communication on the part of medical students in German-speaking countries. The implementation involved very complex technical processes, but went quite smoothly. As expected, there was a correlation between the result of the video exam and that of the practical OSCE exam, which was actually quite low. Consideration of the connection solely between the items regarding procedural knowledge

in the VSE (according to the theoretical concept) and the OSCE turned out to be almost identical in terms of quantity, and it can be assumed that the wider range of topics in the VSE than in the OSCE can explain this result. Presumably, the variety of topics in the VSE overlaps with the classification according to competence levels (see also comments below). With regard to the complexity of the construct of communicative competence and the competence levels presented, it can also be assumed that, in addition to differences in the variety of topics, other areas or other competence levels (apart from factual and procedural knowledge) are also covered by communicative competence in the OSCE. Presumably this is also reflected in the level of the correlation. As expected, the majority of students rated the practical relevance of the examination with the use of videos higher than in the case of purely written examinations, and the doctor-patient consultations were experienced as realistic. Overall, these results suggest that the higher practical relevance that we desired was achieved by comparison with a purely text-based examination. However, it should be noted that the students had no experience with a purely written form of examination in this subject area (and little practical experience in the second semester), which is why their judgement can only be meaningful to a limited extent. A survey of experts would also be constructive in future. In addition, it should be noted that a comparison with a purely written examination was not

Table 4: Evaluation of results

Item	SS18				WS18/19			
	N	Mdn	IQR	yes	N	Mdn	IQR	yes
In your opinion, should the VSE be retained?	142			78%	153			74%
Overall grade of the VSE (School grading system 1-6)	143	2	1		138	3	1	
In the video-based testing of communication skills, the videos generated a higher practical relevance of the content than written exams without video.	143	6	1		138	5	2	
In the video-based testing of communication skills, the videos established a good connection with the practice of medical consultation.	143	6	1		138	5	2	

*Comment.* A total of N=147 participants from the evaluation of SS18 and N=192 participants from WS18/19. Items 3 and 4 in the Table were rated with a seven-point Likert scale from 1 ("totally disagree") to 7 ("completely agree").

implemented in order to avoid unreasonable stress for the students. However, such a comparison would be important in order to determine the greater practical relevance of the VSE than that of purely written examinations. Furthermore, the VSE does not yet adequately reflect the OSCE in its current form.

With regard to the test quality of the VSEs, we believe that the objectivity of their implementation and evaluation can be rated as very good (as expected), since the tasks and the conversation sequences in this test format are standardised and the evaluation is controlled with software. The objectivity can presumably be assessed as higher than that of the OSCE, in which the raters' assessments often show slightly less agreement [17]. However, this assessment is based on plausibility and is not supported by data.

The internal consistencies of the two VSEs and the OSCE turned out to be low, although, contrary to expectations, the internal consistencies of the VSEs were little or no higher than those of the OSCE. This result should be interpreted in the context of the respective examination content. The OSCE that was used consisted of five stations with relatively narrowly defined subject areas. However, the VSEs were designed to cover a greater extent of the learning content, including theoretical concepts (see also Constructive Alignment [32]). Taking the entire course content into account, we found that 30 items are too few to achieve high internal consistency. An increase in the internal consistency of the VSEs could be achieved by increasing the number of items in the subject areas or, as with the OSCE, by reducing the number of subject areas. In accordance with Schecker [33], it seems important to decide whether to focus on bandwidth or specificity when constructing a test. He argued that in order to achieve a consistent scale, it is easier to include items from a single context, but with regard to teaching methodology, a wider range would be more desirable in a complex subject area. With regard to future video-based examinations, this question should therefore be asked and discussed again. According to Schecker, a lower value of internal consistency at the level of the two existing values of the VSEs is well tolerated if the didactic decision is in favour of bandwidth.

At this point, we should discuss the results of the study of the VSEs' factor structure. In both tests, based on the statistics, there is at most a tendency towards one block of topics with regard to patient orientation and empathy,

although many items do not count towards this factor and there is an overall overlap across all items. However, if one looks at the course content, this result is easy to understand. On the one hand, several models and fields of application of the course take into account the topics of patient orientation and empathy. On the other hand, according to the presentation of competence with medical communication as a complex construct, it is evident that different background knowledge and different aspects influence the items. It is plausible that some aspects relate to one another or are equally important in a task. If the results of the internal consistencies are combined with the results of the Principal Axis Analysis, a picture of mostly heterogeneous items is confirmed. A two-factor solution with factual knowledge and procedural knowledge was not in accordance with the results presented. It can be assumed that the effect of mostly heterogeneous subject areas is superimposed on a classification with regard to different levels of competence. The observation that in the VSE in WS18/19, significantly more items load one factor at a high level can probably be explained by the fact that the changed focus of the second exam was on procedural knowledge or tasks regarding the conduct of consultations, and thus reflects patient orientation/empathy. With regard to this point, it is worth reflecting critically on whether it is desirable to prioritise this topic in future exams. In general, it can be assumed that VSEs represent the learning content of the course and thus cover various aspects of communication competence (such as expressing empathy, informing, structuring, using the SPIKES protocol, etc.) as a correspondingly complex construct.

The statistical parameters also show that the VSEs were generally rather easy and the items show a low degree of item-total correlation. A goal for future exams of this kind will therefore be to increase their difficulty so that one can make a better differentiation when assessing skills.

The new form of examination was well received by the students, and the majority of the students supported continuation of the project. It must be noted, however, that the reasons for this have not been recorded and it is possible, for example, that the students have spoken out in favour of retaining the VSE compared to the OSCE due to lower emotional and time-related efforts of the VSE. This needs to be checked again.

While the development of the VSE was initially resource-intensive (as expected due to the large number of videos produced), examination material is now available for a large number of future exams. In our opinion, this means that the performance of the students can be checked with a VSE with fewer resources than with an OSCE (although this statement is not supported by data).

## 5. Conclusion

Finally, if in practice one needs to carry out a test based on the observable level of action, the OSCE still comes closest to the conditions in a real doctor-patient consultation. If an examination needs higher practical relevance than a purely written examination, a video-based examination may well be a suitable form of examination that is more resource-efficient in the medium and long term than an OSCE, and can probably achieve a higher level of objectivity. However, these conclusions are based partly on plausibility (such as resource expenditure and objectivity) and are not supported by data. The internal consistency for both tests may also be influenced by the range of topics and the number of items or stations, while an increase in internal consistency appears to be easier to implement with the VSE than with the OSCE. This also means that it is easier to cover more learning objectives comprehensively with a VSE.

The newly designed video-based e-exam is therefore a very promising instrument for testing certain aspects of the communication competency of medical students, subject to the above-mentioned requirements and limitations. However, there is still a need for improvement in the difficulty of the items, critical reflection on the level of internal consistency, and direct comparison with a purely written examination. Further experience and research at other universities, as well as the development and testing of modified concepts seems necessary and worthwhile with regard to the above-mentioned limitations.

## Notes

The hypotheses and results are based on the dissertation of Ms. Stephanie Ludwig.

## Acknowledgements

We thank the BMBF for funding the project. We would also like to thank Prof. Wermuth, Prof. Hardt and Prof. Beutel, as well as Dr. Schappert, Dr. Ditter and Dr. Seifert, all examiners, simulated patients, scientific assistants, the learning clinic with Mr. Thomas Nowak, the media centre, the ZDV and other participants for their support in implementing this project.

## Funding

This work was supported by the Federal Ministry of Education and Research (Grant number 16DHL1032).

## Profiles

**Location:** Mainz University Medical Center

**Subject/ professional group:** Human medicine

**Number of students per semester:** approx. 200

**Has a longitudinal communication curriculum been implemented?** A pilot implementation is currently underway as part of the LONGKOM project, [<https://www.uni-medizin-mainz.de/lernklinik/startseite/projekte/longkommunikative-kompetenzen-von-aerztinnen-und-aerzten.html>]

**In which semesters are communicative and social skills taught?** 2<sup>nd</sup> and 5<sup>th</sup> semesters, possibly more, currently recorded as part of the LONGKOM project

**Which teaching formats are used?** Role play, simulated patient interviews, lecture, blended learning, sample videos

**In which semesters are communicative and social skills tested (formative or summative and/or graded)?** 2<sup>nd</sup>, 5<sup>th</sup>, 9<sup>th</sup> semester

**Which examination formats are used?** OSCE (summative), video-based single-choice exam (summative), video-based e-exercise (formative), essay exam

**Who (e.g. clinic, institution) is entrusted with development and implementation?** Clinic and Polyclinic for Psychosomatic Medicine and Psychotherapy, Department of Medical Psychology and Medical Sociology, Rudolf Frey Learning Clinic Mainz and practice of Dr. B. Schappert, Mainz

## Current professional roles of the authors

- Dipl.-Psych. Stephanie Ludwig: Research assistant (Clinic and Polyclinic for Psychosomatic Medicine and Psychotherapy, University Medical Center Mainz) in the BMBF-funded 'digiRole' project. Also active as a licensed psychological psychotherapist.
- Lina Behling, MA: Lina Behling is a research assistant at the Center for Quality Assurance and Development at Johannes Gutenberg University, Mainz.
- Univ.-Prof. Dr. Uwe Schmidt: Uwe Schmidt is Professor of University Research at the Institute for Sociology at Johannes Gutenberg University Mainz, and Head of the Center for Quality Assurance and Development. His research focuses on empirical university research and evaluation research.
- Dr. rer. physiol. Dipl.-Psych. Sabine Fischbeck, MME: Research assistant and teaching officer for medical psychology and medical sociology in Mainz, Master of Medical Education and psycho-oncologist.



## Competing interests

The authors declare that they have no competing interests.

## References

- Härtl A, Bachmann C, Blum K, Höfer S, Peters T, Preusche I, Raski B, Rüttermann S, Wagner-Menghin M, Wunsch A, Kiessling C; GMA-Ausschuss Kommunikative und Soziale Kompetenzen. Desire and reality—teaching and assessing communicative competencies in undergraduate medical education in german-speaking europe—a survey. *GMS Z Med Ausbild.* 2015;32(5):Doc56. DOI: 10.3205/zma000998
- Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neill P. Evaluations of situational judgement tests to assess non-academic attributes in selection. *Med Educ.* 2012;46(9):850-868. DOI: 10.1111/j.1365-2923.2012.04336.x
- Fischbeck S, Mauch M, Leschnik E, Beutel ME, Laubach W. Überprüfung ärztlicher kommunikativer Kompetenz mittels einer OSCE bei Studierenden der Medizin im ersten Studienjahr. *Psychother Psychosom Med Psychol.* 2011;61(11):465-471. DOI: 10.1055/s-0031-1291277
- Bundesministerium für Gesundheit. Beschlusstext zum "Masterplan Medizinstudium 2020". Berlin: Bundesministerium für Gesundheit; 2017. Zugänglich unter/available from: [https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/4\\_Pressemitteilungen/2017/2017\\_1/170331\\_Masterplan\\_Beschlusstext.pdf](https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/4_Pressemitteilungen/2017/2017_1/170331_Masterplan_Beschlusstext.pdf)
- Association of American Medical Colleges. Report I. Learning Objectives for Medical Student Education. Guidelines for Medical Schools. Washington: Association of American Medical Colleges; 1998. Zugänglich unter/available from: <https://www.aamc.org/system/files/c/2/492708-learningobjectivesformedicalstudenteducation.pdf>
- UK Foundation Programme Office. The Foundation Programme Curriculum 2016. Birmingham: UK Foundation Programme Office; 2016. Zugänglich unter/available from: [https://www.foundationprogramme.nhs.uk/sites/default/files/2018-07/FP\\_Curriculum\\_2016\\_V2%20\(1\)\\_0.pdf](https://www.foundationprogramme.nhs.uk/sites/default/files/2018-07/FP_Curriculum_2016_V2%20(1)_0.pdf)
- Aspegren K. Beme guide no. 2: Teaching and learning communication skills in medicine—a review with quality grading of articles. *Med Teach.* 1999;21(6):563-570. DOI: 10.1080/01421599978979
- Baribeau DA, Mukovozov I, Sabljic T, Eva KW, deLottinville CB. Using an objective structured video exam to identify differential understanding of aspects of communication skills. *Med Teach.* 2012;34(4):e242-250. DOI: 10.3109/0142159X.2012.660213
- Laidlaw A, Salisbury H, Doherty EM, Wiskin C. National survey of clinical communication assessment in medical education in the united kingdom (UK). *BMC Med Educ.* 2014;14:10. DOI: 10.1186/1472-6920-14-10
- Chong L, Taylor S, Haywood M, Adelstein BA, Shulruf B. Examiner seniority and experience are associated with bias when scoring communication, but not examination, skills in objective structured clinical examinations in australia. *J Educ Eval Health Prof.* 2018;15:17. DOI: 10.3352/jeehp.2018.15.17
- Daniels VJ, Harley D. The effect on reliability and sensitivity to level of training of combining analytic and holistic rating scales for assessing communication skills in an internal medicine resident osce. *Patient Educ Couns.* 2017;100(7):1382-1386. DOI: 10.1016/j.pec.2017.02.014
- Dong T, LaRochelle JS, Durning SJ, Saguil A, Swygert K, Artino AR. Longitudinal effects of medical students' communication skills on future performance. *Mil Med.* 2015;180(4 Suppl):24-30. DOI: 10.7205/MILMED-D-14-00565
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9 Suppl):63-67. DOI: 10.1097/00001888-199009000-00045
- Kopp V, Möltner A, Fischer MR. Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. *GMS Z Med Ausbild.* 2006;23(3):Doc50. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000269.shtml>
- Schubert S, Ortwein H, Dumitsch A, Schwantes U, Wilhelm O, Kiessling C. A situational judgement test of professional behaviour: Development and validation. *Med Teach.* 2008;30(5):528-533. DOI: 10.1080/01421590801952994
- Koczwara A, Patterson F, Zibarras L, Kerrin M, Irish B, Wilkinson M. Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Med Educ.* 2012;46(4):399-408. DOI: 10.1111/j.1365-2923.2011.04195.x
- Chenot JF, Ehrhardt M. Objective structured clinical examination (OSCE) in der medizinischen Ausbildung: Eine Alternative zur Klausur. *Z Allg Med.* 2003;79(9):437-442. DOI: 10.1055/s-2003-43064
- Schmitz FM, Schnabel KP, Bauer D, Bachmann C, Woermann U, Guttormsen S. The learning effects of different presentations of worked examples on medical students' breaking-bad-news skills: A randomized and blinded field trial. *Patient Educ Couns.* 2018;101(8):1439-1451. DOI: 10.1016/j.pec.2018.02.013
- Humphris GM, Kaney S. The objective structured video exam for assessment of communication skills. *Med Educ.* 2000;34(11):939-945. DOI: 10.1046/j.1365-2923.2000.00792.x
- Hulsman RL, Mollema ED, Oort FJ, Hoos AM, de Haes JC. Using standardized video cases for assessment of medical communication skills: Reliability of an objective structured video examination by computer. *Patient Educ Couns.* 2006;60(1):24-31. DOI: 10.1016/j.pec.2004.11.010
- Fischbeck S, Mauch M, Unterrainer J. Vom Kompetenzwissen zum praktischen Können: Arzt-Patient-Kommunikation prüfen mit Key-Feature-Klausuren und OSCE. In: Berth H, editor. *Abstracts zur Jahrestagung der Deutschen Gesellschaft für Medizinische Psychologie 2013.* Lengerich: Pabst-Publishers; 2013. p.126-127.
- Kurtz S, Silverman J, Draper J. *Teaching and learning communication skills in medicine.* 2nd ed. Oxford: Radcliffe; 2006.
- Back AL, Arnold RM, Baile WF, Fryer-Edwards KA, Alexander SC, Barley GE, Gooley TA, Tulsy JA. Efficacy of communication skills training for giving bad news and discussing transitions to palliative care. *Arch Intern Med.* 2007;167(5):453-460. DOI: 10.1001/archinte.167.5.453
- Kanfer FH, Reinecker H, Schmelzer D. *Selbstmanagement-Therapie.* 5. Aufl. Berlin, Heidelberg: Springer; 2012. DOI: 10.1007/978-3-642-19366-8
- Hamann J, Loh A, Kasper J, Neuner B, Spies C, Kissling W, Härter M, Heesen C. Partizipative Entscheidungsfindung. Implikationen des "shared decision making" für Psychiatrie und Neurologie. *Nervenarzt.* 2006;77(9):1071-1078. DOI: 10.1007/s00115-005-1950-5
- Baile WF, Buckman R, Lenzi R, Glober G, Beale EA, Kudelka AP. Spikes—a six-step protocol for delivering bad news: Application to the patient with cancer. *Oncologist.* 2000;5(4):302-311. DOI: 10.1634/theoncologist.5-4-302

27. Hoppe-Seyler T, Gartmeier M, Möller G, Bauer J, Wiesbeck A, Karsten G. Entwicklung von Lehrfilmen zur Gesprächsführung zwischen Realitätsnähe und systematischer didaktischer Gestaltung. ZFHE. 2014;9(3):127-135. DOI: 10.3217/zfhe-9-03/14
28. Edwards AW. The Measure of Association in a  $2 \times 2$  Table. J Royal Stat Soc. 1963;126(1):109-114. DOI: 10.2307/2982448
29. Leonhart R. Lehrbuch Statistik. 2. Aufl. Bern: Huber; 2009.
30. Wermuth N. How can graphical Markov models aid causal inference? In: DAGStat 2019. Statistics under one umbrella. Conference Guide. München, 18.-20. März 2019. Göttingen: Deutsche Arbeitsgemeinschaft Statistik; 2019.
31. Rost DH, Schermer FJ. Strategien der Prüfungsangstverarbeitung. Z Diff D P. 1986;7(3):127-139.
32. Biggs J. Enhancing teaching through constructive alignment. High Educ. 1996;32(3):347-364. DOI: 10.1007/BF00138871
33. Schecker H. Überprüfung der Konsistenz von Itemgruppen mit Cronbachs alpha. In: Krüger D, Parchmann I, Schecker H, editors. Methoden in der naturwissenschaftsdidaktischen Forschung. Berlin: Springer Spektrum; 2014. chapter online-Zusatzmaterial. DOI: 10.1007/978-3-642-37827-0\_1

**Corresponding author:**

Dipl.-Psych. Stephanie Ludwig  
University Medical Center, Johannes Gutenberg University  
Mainz, Clinic and Polyclinic for Psychosomatic Medicine  
and Psychotherapy, Department of Medical Psychology  
and Medical Sociology, Saarstr. 21 (Campus Universität),  
D-55099 Mainz, Germany  
stephsch@uni-mainz.de

**Please cite as**

Ludwig S, Behling L, Schmidt U, Fischbeck S. Development and testing of a summative video-based e-examination in relation to an OSCE for measuring communication-related factual and procedural knowledge of medical students. *GMS J Med Educ.* 2021;38(3):Doc70. DOI: 10.3205/zma001466, URN: urn:nbn:de:0183-zma0014667

**This article is freely available from**

<https://www.egms.de/en/journals/zma/2021-38/zma001466.shtml>

**Received:** 2020-01-29

**Revised:** 2020-10-25

**Accepted:** 2020-11-20

**Published:** 2021-03-15

**Copyright**

©2021 Ludwig 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 und Erprobung einer summativen videobasierten E-Prüfung mit Bezug zu einer OSCE zur Messung von kommunikationsbezogenem Fakten- und Handlungswissen bei Medizinstudierenden

## Zusammenfassung

**Zielsetzung:** Die Überprüfung ärztlicher kommunikativer Kompetenzen erfolgt im Rahmen des Medizinstudiums primär über die Prüfungsform OSCE [1]. Diese ermöglicht eine Bewertung praktischer Leistungen, ist allerdings ressourcenaufwändig und bringt Limitationen hinsichtlich der Testgüte mit sich. Die Zielsetzung des vom BMBF geförderten Projekts „Digitales Prüfungstool zur Messung kommunikativer Fertigkeiten im Medizinstudium“ (digiRole) war daher die Entwicklung digitaler Formate von Prüfungen kommunikativer Kompetenzen von Medizinstudierenden als elektronisch gestützte Varianten einer OSCE. Die entsprechenden Prüfungsformen sollten dabei ökonomisch sein, einen hohen Praxisbezug herstellen und eine hohe psychometrische Qualität aufweisen. Inhaltlich sollten die Prüfungsfragen Fakten- und Handlungswissen als Bestandteile der kommunikativen Kompetenz abbilden, wobei davon ausgegangen wurde, dass Handlungswissen stärker mit der OSCE-Leistung zusammenhängt als Faktenwissen. Der vorliegende Artikel beschreibt die Entwicklung und Erprobung einer bestehensrelevanten videobasierten kommunikationsbezogenen E-Prüfung als ersten Meilenstein des Gesamtprojekts.

**Methodik:** Auf Grundlage der Lerninhalte eines vorklinischen Kurses zur Arzt-Patient-Kommunikation der Universitätsmedizin Mainz im Fach Medizinische Psychologie und Medizinische Soziologie wurden Videos produziert und dazu in Bezug stehende Prüfungsfragen mit Aufgaben im Sinne eines Situational Judgement Tests [2] konzipiert. Im Sommersemester 2018 nahmen 226 Studierende an dieser videobasierten Single-Choice Prüfung (Video-based Single-Choice Examination, VSE) und im Wintersemester 2018/2019 192 Studierende (unterschiedliche Kohorten) an der VSE sowie an einer bereits gut erprobten Kommunikations-OSCE mit fünf Stationen teil [3].

**Ergebnisse:** Die internen Konsistenzen lagen für die VSE im SS 2018 bei  $\alpha=.55$ , im WS 2018/19 bei  $\alpha=.62$  und für die OSCE bei  $\alpha=.60$ . Die Leistungen der Studierenden in VSE und OSCE waren positiv korreliert ( $r=.21$ ,  $p\leq.01$ ). Hauptachsenanalysen ergaben keine Dimensionierung in die Aspekte Fakten- und Handlungswissen. Die Studierenden gaben in der Evaluation mehrheitlich an, dass in der VSE der Bezug zur Praxis des ärztlichen Gesprächs gut hergestellt wird und sprachen sich für eine Beibehaltung dieser Prüfungsform aus.

**Schlussfolgerung:** Der Zusammenhang zwischen VSE und OSCE fällt relativ gering aus, so dass die VSE in dieser Form noch nicht hinreichend ein OSCE-Ergebnis präzisieren kann. Hinsichtlich der internen Konsistenz konnten die VSE und die OSCE ein nahezu identisches Ergebnis erzielen. Es kann des Weiteren angenommen werden, dass die VSE durch den Einsatz von standardisierten videobasierten Prüfungsaufgaben einen hohen Grad der Objektivität sowie ein im Vergleich zu OSCEs ressourcenschonenderes Vorgehen erzielen kann.

Stephanie Ludwig<sup>1</sup>

Lina Behling<sup>2</sup>

Uwe Schmidt<sup>2</sup>

Sabine Fischbeck<sup>1</sup>

1 Universitätsmedizin der Johannes Gutenberg-Universität Mainz, Klinik und Poliklinik für Psychosomatische Medizin und Psychotherapie, Schwerpunkt Medizinische Psychologie und Medizinische Soziologie, Mainz, Deutschland

2 Johannes Gutenberg-Universität Mainz, Zentrum für Qualitätssicherung und -entwicklung (ZQ), Mainz, Deutschland

**Schlüsselwörter:** Medizinische Ausbildung, kommunikative Kompetenzen, summative Beurteilung, videobasierte Untersuchung, OSCE

## 1. Einleitung

### 1.1. Forschungsstand

Im Rahmen des Medizinstudiums gewinnt die Lehre und Überprüfung kommunikativer Kompetenzen zunehmend an Relevanz. Der „Masterplan Medizinstudium 2020“ betont die Bedeutung entsprechender Ausbildung im Rahmen des Medizinstudiums aktuell erneut [4]. Auch international ist die Ausbildung kommunikativer Kompetenzen ein wesentlicher Teil der medizinischen Ausbildung [5], [6]. Hinsichtlich der praktizierten Formen von Lehre und Prüfung im Bereich kommunikative ärztliche Kompetenz konnten Härtl et al. [1] in ihrer Umfrage bezogen auf Hochschulen im deutschsprachigen Raum ermitteln, dass sich in 70% der Studiengänge die Lehrenden an Modellen oder Lernzielkatalogen orientieren. Das Konstrukt ärztliche kommunikative Kompetenz gilt dabei als komplex und vielschichtig [7], [8]. Die Übersichtsarbeit [1] für den deutschsprachigen Raum sowie eine solche von Laidlaw et al. [9] für Großbritannien zeigen zudem hinsichtlich der Prüfungsformen auf, dass die Überprüfung kommunikativer Kompetenzen vor allem in Form der praktischen Prüfungsform OSCE (Objective Structured Clinical Examination) erfolgt. Im deutschsprachigen Raum ist die schriftliche Prüfung zudem die zweithäufigste Form. Auch in anderen Ländern wie Australien, Kanada und den USA zeigt sich, dass neben dem Einsatz von klassischen schriftlichen Prüfungen die Prüfungsform OSCE Anwendung findet [10], [11], [12]. Bei der OSCE wird anhand simulierter Situationen an mehreren Prüfungsstationen teilweise unter Einsatz von Simulationspatienten die Kompetenz der Studierenden meist anhand von Checklisten durch die Prüfer eingeschätzt. Bei einem Vergleich schriftlicher Prüfungsformen mit der OSCE besteht der Vorteil einer OSCE in dem hohen Praxisbezug und der Möglichkeit einer Leistungsprüfung auf verhaltensbezogener Ebene [13]. Demgegenüber bewegen sich schriftliche Prüfungen auf der kognitiven Ebene, die sich in Faktenwissen und Handlungswissen gliedern lässt [13]. Diese Ebene ist beispielsweise über Single/Multiple Choice Fragen, Key-Feature Konzepte (Treffen kritischer Entscheidungen) bzw. Situational Judgement Tests prüfbar [2], [14], [15], [16]. Bei einem Situational Judgement Test werden schriftlich oder videobasiert Szenarien aus der Berufspraxis präsentiert, zu denen sich im Rahmen von Prüfungsfragen für eine Handlungsoption entschieden werden soll. Das Antwortformat kann unterschiedlich ausgestaltet werden. Die internen Konsistenzen sind dabei zumeist auf einem guten Niveau [2]. In der medizinischen Ausbildung kamen Situational Judgment Tests bereits zum Einsatz [15]. In Bezug auf die Ökonomie be-

deutet die Durchführung einer OSCE einen hohen Zeitaufwand und gestaltet sich auch sehr ressourcenintensiv hinsichtlich finanzieller und personeller Aspekte. Die Reliabilität und Inhaltsvalidität einer OSCE wird maßgeblich von der Anzahl der Stationen bestimmt [17], wodurch der Aufwand immens und oft in der anzustrebenden Form institutionell nicht leistbar ist. Die Bewertung durch verschiedene Prüfer kann unterschiedlich ausfallen, womit die Interrater-Reliabilität und damit die Objektivität in einigen Fällen unzureichend ausfällt [17]. Eine schriftliche Prüfung wie der Situational Judgement Test kann hingegen durch die Standardisierung von Fragen und Auswertung sowie mehr Fragen in der Regel eine höhere Objektivität und Reliabilität erzielen.

Die dargelegten Vor- und Nachteile der beiden häufigsten Prüfungsformen ärztlicher kommunikativer Kompetenz führten zu der Frage nach alternativen Prüfungsformen, die sowohl der Relevanz eines hohen Praxisbezugs als auch den Anforderungen an ein ressourcenschonendes Vorgehen und eine hohe Testgüte Rechnung tragen. Der Einsatz digitaler Formate mit Videos erscheint im Themenfeld kommunikativer Kompetenzen optimal, um mit hohem Praxisbezug und bei hohen Studierendenanzahlen fallbasiert prüfen zu können. Es entspricht studentischen Rückmeldungen aus dem Unterricht, dass in Videos dargestellte Arzt-Patient-Gespräche in ihrer Darstellung sehr viel näher an eine reale Situation heranreichen als nur schriftlich formulierte Fallvignetten, bei denen z.B. Bild, Ton, Mimik und Gestik komplett fehlen. Studierende können sich so ggf. besser in die Rolle des Arztes hineinversetzen und „erleben“ die Situation auch eher als bei rein schriftlichen Prüfungen. Zudem bietet sich die Möglichkeit der Einbindung von Fragen und Videos in Prüfungssoftware, wodurch die Prüfung ökonomisch und objektiv gestaltet werden kann. Ein Einsatz digitaler Formate im Bereich der Lehre ärztlicher kommunikativer Kompetenzen ist zum Zeitpunkt des Projektbeginns 2017 im deutschsprachigen Raum den Autoren hauptsächlich in Bezug auf videogestützte Übungen zum Erlernen [18] kommunikativer Kompetenzen bekannt. Für den Bereich der Prüfungen ergibt sich lediglich bei Härtl et al. [1] ein Hinweis darauf, dass an zwei Universitäten Videos in Prüfungen bereits zum Einsatz kamen. International konnten mehrere relevante Studien recherchiert werden, die sich mit der Entwicklung und Erprobung videobasierter Prüfungen zur Messung ärztlicher kommunikativer Kompetenzen befassen [8], [19], [20]. An den Universitäten der vorgenannten Studien kam jeweils eine sogenannte OSVE (Objective Structured Video Exam) zum Einsatz, bei der die Studierenden Aufgaben (u.a. Fragen im Short-Essay Format, Benennung von kommunikativen Fertigkeiten) in Bezug auf Videos zu Arzt-Patient-Gesprächen bearbeiten mussten. Bei Humphris und Kaney [19] ergab sich eine Korrelation von  $r=.17$  zwischen OSVE und OSCE.



## 1.2. Projektdesign

Der dargestellte Mangel an ressourcenschonenden kompetenzorientierten Prüfungen mit hohem Praxisbezug leitete uns somit zu der Idee erstmalig für den deutschsprachigen Raum eine videobasierte Prüfung von Fakten- und Handlungswissen als Bestandteile kommunikativer Kompetenzen im Medizinstudium zu entwickeln, zu erproben und wissenschaftlich zu untersuchen. Die Prüfung sollte dabei als digitale Variante einer OSCE angelegt werden. In einer ersten Studienphase sollte die videobasierte Prüfung mit Aufgaben im Sinne eines Situational Judgement Tests als bestehensrelevante Prüfung (Video-based Single-Choice Examination, VSE) ausgestaltet werden. Wir gehen davon aus, dass die VSE aufgrund oben ausgeführter Gründe einen höheren Praxisbezug als rein schriftliche Prüfungen erzielen kann (u. a. eingeschätzt durch die Studierenden), sich als ökonomisch erweist und eine hohe Durchführungs- und Auswertungsobjektivität (durch Softwaresteuerung) erreicht. Die hohe Ökonomie ist dadurch zu erwarten, dass zwar vermutlich zunächst ein hoher zeitlicher und finanzieller Aufwand für die Erstellung der Szenarien und die Produktion der Videos betrieben werden muss, dann aber mutmaßlich ausreichend Material vorliegt, um kostengünstig in mehreren Semestern Studierende prüfen zu können. Gegenüber der OSCE könnte so ggf. ein hoher Zeitaufwand für Prüfer minimiert werden. Zudem wird angenommen, dass zwischen den Prüfungsleistungen der Studierenden in VSE und OSCE ein positiver Zusammenhang besteht (vgl. auch [19] sowie [21] mit  $r = .32$  Korrelation von Key-Feature-Prüfung und OSCE). Die zugrundeliegenden Dimensionen der VSEs sollen untersucht werden, wobei es denkbar ist, dass sich die Konzeption der Prüfung entsprechend der beiden Aspekte „Faktenwissen“ und „Handlungswissen“ in einer zweifaktoriellen Lösung über alle inhaltlichen Themen hinweg widerspiegelt. Sollte sich eine solche Lösung ergeben, soll des Weiteren untersucht werden, ob die Fragen in der VSE auf der Ebene des Handlungswissens in einem höheren Zusammenhang mit den Ergebnissen der OSCE stehen als die Fragen auf der Ebene des Faktenwissens, da augenscheinlich Handlungswissen und die durchgeführte Handlung stärker miteinander zusammenzuhängen scheinen als Faktenwissen und die durchgeführte Handlung. Alle Prüfungen sollten durch die Studierenden evaluiert werden, um den Praxisbezug sowie die Frage nach Beibehaltung dieser Prüfungsform bewerten zu lassen.

## 2. Methode

### 2.1. Studienteilnehmer und -ablauf

Die VSE kam im Kursus II „Arzt-Patient-Kommunikation“ bei den Studierenden des Medizinstudiums im zweiten Semester an der Universitätsmedizin Mainz im SS 2018 erstmals zum Einsatz. Im WS 2018/19 absolvierten die Studierenden im zweiten Semester der nachfolgenden

Kohorte im vorgenannten Kurs ebenfalls die VSE und darüber hinaus die OSCE. Im Vorfeld der Durchführung gab es in den Kursen Ausführungen zur genauen Ausgestaltung der Prüfungen mit einer Beispielaufgabe. Zudem gab es ein ausführliches Informationsblatt zum eigenständigen Nachlesen. Die VSEs wurden in den Räumlichkeiten des Zentrums für Datenverarbeitung der Universität Mainz zu Semesterende durchgeführt. Die technische Umsetzung erfolgte mit der Prüfungssoftware ILIAS. Bis auf vereinzelte Probleme mit der Tonwiedergabe und den Ladungszeiten der Videos, die jeweils schnell und gut behoben werden konnten, verlief die Durchführung der VSEs unproblematisch und reibungslos. Nach Prüfungsende wurde jeweils die Evaluation anhand von Paper-Pencil-Befragungen durchgeführt. Die höhere Rücklaufquote bei der Evaluation der VSE in der zweiten Kohorte ist vermutlich darauf zurückzuführen, dass die Prüfungsaufsichten aufgrund der Erfahrungen aus dem vorherigen Semester stärker auf die Evaluation hingewiesen und diese auch gezielt eingesammelt haben. Angaben zu Anzahl, Alter und Geschlecht der Studienteilnehmer sind in Tabelle 1 aufgeführt.

### 2.2. Entwicklung VSE

In der Entwicklungsphase der VSE wurden angelehnt an die Inhalte der Prüfungen vergangener Jahre in einem ersten Schritt Kasuistiken und Dialoge zwischen Arzt und Patient entwickelt. Die Dialoge beinhalten die wesentlichen theoretischen Konzepte, die Studierende in der Praxis anwenden können sollen (u.a. Calgary-Cambridge Schema [22], NURSE-Modell [23], Gesprächsführungstechniken wie aktives Zuhören [24], Prinzip der Partizipativen Entscheidungsfindung [25], SPIKES-Protokoll [26]). Die Szenarien beziehen sich auf folgende Erfordernisse: Medikamentencompliance prüfen und fördern, Anamnesegespräch führen, Krebsdiagnose nach SPIKES-Protokoll mitteilen, Visitingespräch führen, mit Patient Therapie bei Hypertonie nach dem Prinzip der Partizipativen Entscheidungsfindung absprechen, bei Patientin mit Adipositas eine Verhaltensänderung herbeiführen, eine Analyse der Stressreaktion vornehmen und Informationen vermitteln. Die erstellten Dialoge wurden von Kursdozenten überprüft und überarbeitet. Zur Gewährleistung der Authentizität erfolgte zudem eine Expertenbefragung (u.a. Allgemeinmediziner und Psychologen). Die Experten wurden danach ausgewählt, dass sie über praktische Erfahrungen im Berufsalltag verfügen, teilweise zusätzlich selbst lehren oder in einem Fall bereits an der Entwicklung einer früheren OSCE beteiligt waren. Die Experten gaben schriftlich (orientiert an gestellten Fragen, beispielsweise Bitte um Einschätzung der Glaubwürdigkeit von Setting, Krankengeschichte, Gesprächsablauf, Patientendarstellung u.s.w.) ihre Rückmeldung zu den Dialogen ab, die qualitativ ausgewertet und auf deren Basis die Dialoge entsprechend überarbeitet wurden.

Die Produktion der Videos wurde in Zusammenarbeit mit dem Medienzentrum (Universität Mainz) und der Lernklinik (Universitätsmedizin Mainz) realisiert. Simulationspa-

Tabelle 1: Studienteilnehmer

	VSE SS18	VSE WS18/19	OSCE WS18/19	OSCE und VSE WS18/19
<b>N</b>	226	193	195	192
<b>Alter</b>	M = 23 SD = 4.3 (18-45)	M = 25 SD = 3.8 (19-49)*	M = 25 SD = 3.8 (19-49)*	M = 25 SD = 3.8 (19-49)*
<b>Geschlecht weiblich</b>	58%	58%	57%	58%
<b>Geschlecht männlich</b>	42%	42%	43%	42%

Anmerkung. \*2 missing.

tienten sowie ein Allgemeinmediziner übernahmen die Rollen. In der Bearbeitung des Drehmaterials wurde ein Split-Screen-Verfahren umgesetzt (jeweils Totale und Frontalaufnahmen), damit Hilfestellungen durch Perspektivwechsel vermieden werden können. Entsprechend der Erfahrungen von Hoppe-Seyler et al. [27], wie bedeutsam die von Studierenden eingeschätzte Glaubwürdigkeit der dargestellten Szenen ist, erfolgte im Vorfeld der Hauptdreharbeiten eine Probeproduktion zu einer Szene, welche von acht Studierenden (über einen E-Mailverteiler von Medizinstudierenden verschiedener Semester) evaluiert wurde. Arztrolle (63%), Patientenrolle (75%), Praxis (88%) und Interaktion (88%) wurden dabei von der Mehrzahl als authentisch eingeschätzt.

Die auf die Prüfungsvideos bezogenen Prüfungsfragen wurden im Single-Choice Format mit jeweils fünf Antwortmöglichkeiten und unter Bezugnahme vorhandener Prüfungsfragen videoangepasst entwickelt. In den Aufgaben wird entweder Faktenwissen zu der dargestellten Sequenz abgefragt oder eine Handlungsentscheidung/Auswahl einer geeigneten Formulierung im Sinne eines Situational Judgement Tests eingefordert. Dieses Vorgehen sollte der Kurskonzeption mit theoretischen und praktischen Inhalten Rechnung tragen. Zudem wurde entschieden die vielfältigen Themenfelder, die im Kursus behandelt werden, in ihrer Bandbreite abzufragen und nicht eine Eingrenzung auf bestimmte Themen vorzunehmen. In der ersten Prüfung wurde der Schwerpunkt etwas mehr auf Fakten- und in der zweiten Prüfung etwas mehr auf Handlungswissen gelegt (Aufteilung in Fakten- und Handlungswissen siehe Tabelle 2). Die Fragen wurden von Kursdozenten überprüft und überarbeitet.

Die Prüfungsaufgaben waren im folgenden Stil gehalten: Im Video wird der Fall eines 50-jährigen Patienten gezeigt, bei dem eine Hypertonie diagnostiziert wurde und bei dem nach dem Prinzip der Partizipativen Entscheidungsfindung eine Behandlungsentscheidung getroffen werden soll. Im Video ist zu sehen, wie die Ärztin dem Patienten mitteilt, dass es ihr wichtig sei gemeinsam mit ihm eine Entscheidung über die Behandlungsschritte zu treffen und der Patient teilt mit, dass er dies begrüße. Die sich daran anschließende Aufgabe lautet „Ärztin und Patient haben nun eine Gleichberechtigung hinsichtlich der Entscheidung ausgesprochen. Bei der Umsetzung der Partizipativen Entscheidungsfindung im Arzt-Patient-Gespräch ist eine Reihenfolge bestimmter Schritte zu berücksichtigen.“

Welcher Schritt folgt nun nach dem Modell der Partizipativen Entscheidungsfindung? Und die fünf Antwortmöglichkeiten sind „Mitteilen, dass eine Entscheidung ansteht“, „über Wahlmöglichkeiten informieren“ (richtige Antwort), „Information über Vor- und Nachteile der Optionen geben“, „Verständnis, Gedanken und Erwartungen erfragen“ und „Präferenzen ermitteln“.

Die Auswertung erfolgte über eine Dichotomisierung der fünf Antwortmöglichkeiten („richtig“/„falsch“) mit nachfolgender Ermittlung eines Summenwertes. Entsprechend der weiter unten berichteten Ergebnisse der Hauptachsenanalyse wurden fünf Items aus dem SS18 im WS18/19 erneut eingesetzt.

### 2.3. OSCE

Die OSCE wurde entsprechend langjähriger Erfahrungen mit einer Kommunikations-OSCE in Mainz [3] anhand bereits fünf entwickelter Stationen (Stationen zu den Oberthemen Anamnese, Analyse der Stressreaktion, Compliance überprüfen und fördern, Partizipative Entscheidungsfindung und Mitteilung einer Krebsdiagnose nach SPIKES-Protokoll) geprüft. Anhand von dazu passenden, ebenfalls bereits entwickelten und erprobten Checklisten wurden die Leistungen der Studierenden von insgesamt 11 Prüfern über drei Prüfungstage hinweg beurteilt. Die Checklisten bezogen sich sowohl auf die spezifischen Inhalte der Stationen als auch auf Aspekte der Gesprächsführung wie Empathie, aktives Zuhören, Blickkontakt und offene Körperposition (Item z.B. „nahm offene Körperposition ein“). Die Inhalte decken sich damit in den wesentlichen Themen mit den Inhalten der VSEs, wobei bei diesen noch weitere Themen sowie mehr theoretisches Wissen hinzugenommen wurden. Die Bewertung der Leistungen erfolgte über eine vorab festgelegte Gewichtung der Aspekte der Checklisten, wobei für jede Station max. 7 Punkte und somit insgesamt max. 35 Punkte erzielt werden konnten (absolute Bestehensgrenze 60%).

### 2.4. Statistische Auswertung

Die internen Konsistenzen der Prüfungen wurden jeweils mit Cronbachs  $\alpha$  ermittelt. Anhand des Rangkorrelationskoeffizienten von Spearman wurden die Zusammenhänge zwischen den Leistungen der Studierenden in den Prüfungen VSE und OSCE betrachtet. Kolmogorov-Smirnov-Tests wurden zur Prüfung der Normalverteilung der Variablen eingesetzt. Mit den Aufgaben der VSEs beider Semester

Tabelle 2: Itemcharakteristiken der VSEs im SS18 (N=226) und im WS18/19 (N=193)

VSE SS18				VSE WS18/19			
	richtig	falsch	Trennschärfe $r_{it}$		richtig	falsch	Trennschärfe $r_{it}$
Item (Oberthemen)				Item (Oberthemen)			
<b>Items Handlungswissen</b>							
4 Patientenorientierung/ Empathie (bei Non-Compliance) *	202	24	.30	22 Patientenorientierung/ Empathie (bei Non-Compliance) *	174	19	.06
11 Patientenorientierung/ Empathie *	220	6	.33	23 Patientenorientierung/ Empathie *	189	4	.17
16 Patientenorientierung/ Empathie *	213	13	.31	25 Patientenorientierung/ Empathie *	190	3	.30
19 Patientenorientierung/ Empathie (bei Visite) *	214	12	.27	26 Patientenorientierung/ Empathie (bei Visite) *	181	12	.26
1 Verständliche Informationsvermittlung	212	14	.16	2 Risikokommunikation	89	104	.15
7 Stressanalyse	109	117	.14	3 Patientenorientierung/ Empathie (Positive Wertschätzung)	184	9	.36
9 SPIKES-Protokoll	208	18	.10	5 Patientenorientierung/ Empathie	175	18	.16
12 Partizipative Entscheidungsfindung	159	67	.18	7 Patientenorientierung/ Empathie (Motivational Interviewing)	155	38	.29
15 Verhaltensmodifikation	139	87	.08	8 Verständliche Informationsvermittlung	169	24	.07
20 Visite, Struktur, Informierung	224	2	.14	9 Patientenorientierung/ Empathie (Asymmetrisches Vermeidungsverhalten)	147	46	.32
25 Umgang mit Nebenwirkungen	135	91	.08	10 Partizipative Entscheidungsfindung	32	161	.19
27 Gesprächsförderer	225	1	-.04	14 Anamnese	155	38	.18
30 Verständliche Informationsvermittlung	57	169	-.01	15 Anamnese	166	27	.22
				20 Compliance	179	14	.24
				21 Motivational Interviewing	182	11	.30
				29 Aktives Zuhören	122	71	.24
				30 Patientenorientierung/ Empathie ('schwieriger' Patient)	182	11	.44
<b>Items Faktenwissen</b>							
13 Patientenorientierung/ Empathie (Schulz von Thun) *	219	7	.17	24 Patientenorientierung/ Empathie (Schulz von Thun) *	186	7	.26
2 Anamnese	187	39	.15	1 SPIKES-Protokoll	86	107	.08
3 Fragetypen	169	57	.20	4 Operante Konditionierung	145	48	.26
5 Anamnese	150	76	.14	6 Modell „Stages of Change“	150	43	.04
6 Metakommunikation	100	126	.21	11 Patientenrolle im Allgemeinkrankenhaus	147	46	.26
8 Aktives Zuhören	204	22	-.01	12 Patientenorientierung/ Empathie	179	14	.34
10 Nonverbale Kommunikation	185	41	.26	13 Arzt-Patient-Beziehung	180	13	.24
14 Modell „Stages of Change“	160	66	.18	16 Beurteilungsfehler	185	8	.23
17 Beurteilungsfehler	220	6	.04	17 NURSE-Modell	189	4	.12
18 Partizipative Entscheidungsfindung	187	39	.19	18 Fragetypen	161	32	.14
21 NURSE-Modell	210	16	.09	19 Nonverbale Kommunikation	157	36	.11
22 Asymmetrisches Vermeidungsverhalten	195	31	.21	27 Verhaltensmodifikation	119	74	.05
23 Theorie zu Empathie	20	206	.11	28 Konzept Kontrollüberzeugung	101	92	.04
24 Transaktionales Stressmodell	126	100	.13				
26 Patientenorientierung/ Verhaltensmodifikation	193	33	.26				
28 ICF-Modell der WHO	113	113	.21				
29 Motivational Interviewing	115	111	.12				

Anmerkung. Die mit einem \* markierten Items wurden in beiden Prüfungen eingesetzt. Bei den anderen Items handelt es sich prüfungsübergreifend teilweise zudem um dasselbe Oberthema, allerdings ist die inhaltliche Ausgestaltung der Aufgabe und der Frage verschieden. Richtig und falsch sind Häufigkeitsangaben.

wurde jeweils eine iterative Hauptachsenanalyse durchgeführt. Da es sich bei den betrachteten Items um dichotome Variablen handelt, die zusätzlich Itemschwierigkeiten von teilweise sehr gering oder sehr hoch zeigen (vgl. Abschnitt Ergebnisse und Tabelle 2) und somit ungleiche Randsummenverteilungen vorliegen, wurden Odds Ratios zwischen allen Items berechnet. Odds Ratios weisen Werte zwischen Null und Unendlich auf und wurden daher anhand der von Yules vorgeschlagenen Formel auf Werte zwischen -1 bis +1 transformiert. Die transformierten Odds Ratios bilden die Matrix, auf der die iterative Hauptachsenanalyse beruht. Es wurde Yules Y mit der Formel  $Y = (\sqrt{OR} - 1) / (\sqrt{OR} + 1)$  verwendet [28], [29], [30].

### 3. Ergebnisse

Die Itemcharakteristiken für die beiden VSEs sind in Tabelle 2 aufgeführt. Die Mehrzahl der Items wurde von den meisten Studierenden richtig beantwortet ( $M = .75$  SS18 und  $M = .80$  WS18/19 über alle Mittelwerte). Die Trennschärfen der Items der VSEs liegen in einem Bereich von -.04 bis .44.

Für das Extrahieren der Anzahl relevanter Faktoren wurden die Kriterien von Rost und Schermer [31] herangezogen. Im SS18 ergaben sich 13 Eigenwerte größer 1, wobei anhand des Screeplots nur ein einziger Faktor sinnvoll erschien. In diesem Faktor wurden nur Variablen belassen, die eine Kommunalität von  $h^2 \geq .16$  aufweisen und für deren Absolutbetrag  $a \geq .40$  gilt (vgl. Tabelle 3). Die verbleibenden 8 Items sind in dieser Form kaum eindeutig auf nur einen inhaltlichen Aspekt zu reduzieren. Der inhaltlichen Plausibilität vor der mathematischen Lösung Vorrang einräumend, ergibt allerdings eine Zusammenführung von 5 Items (Item 4, 11, 13, 16, 19) eine deutlich passendere inhaltliche Ausrichtung unter dem Aspekt „patientenorientiertes, einfühlsames Agieren und Verstehen anhand der Situation des Patienten und der Gesprächsinhalte“ (Cronbachs  $\alpha = .62$ ,  $N = 226$ ). Im WS18/19 ergab sich bei 12 Eigenwerten über 1 ebenfalls nur 1 Faktor anhand des Screeplots, wobei 14 Items nach Anwendung der Kriterien verbleiben. Hier können 10 der 14 Items ebenfalls dem vorgenannten Thema zugeordnet werden (3, 5, 7, 9, 12, 23, 24, 25, 26, 30; Cronbachs  $\alpha = .60$ ,  $N = 193$ ). Vier weitere Items weisen inhaltlich Überschneidungen mit dem Thema, aber auch mit anderen Themen auf, sodass sie nicht eindeutig dem Faktor zugeordnet werden können. Insgesamt ist anzumerken, dass die vorgenommenen Zuordnungen bzw. Trennungen unter Beachtung der Inhalte jeweils aller Items nicht vollständig und bis zuletzt scharf sind. Von den 5 Items aus dem SS18 laden 4 Items erneut auf dem Faktor im WS18/19 (Items 23-26 WS18/19).

Die interne Konsistenz der VSE im SS18 lag bei einem Cronbach-Alpha-Koeffizienten von  $\alpha = .55$  ( $N = 226$ ) und im WS18/19 bei  $\alpha = .62$  ( $N = 193$ ). Für die OSCE ergab sich eine interne Konsistenz von  $\alpha = .60$  ( $N = 195$ ).

Der Median der VSE im SS18 lag bei  $Mdn = 23$ ,  $IQR = 3$  für  $N = 226$ , der VSE im WS18/19 bei  $Mdn = 25$ ,  $IQR = 3$  für

$N = 192$  und für die OSCE bei  $Mdn = 26.25$ ,  $IQR = 2.75$  für ebenfalls  $N = 192$ . Der Rangkorrelationskoeffizient nach Spearman in Bezug auf VSE und OSCE im WS18/19 lag bei  $r = .21$  ( $p \leq .01$ ). Die beiden Variablen sind nicht normalverteilt (Kolmogorov-Smirnov-Test OSCE  $D(192) = .09$  und VSE WS18/19  $D(192) = .15$ , jeweils  $p \leq .01$ ). Eine zusätzliche Betrachtung des Zusammenhangs der Items ausschließlich zu Handlungswissen der VSE und dem OSCE Ergebnis ergibt des Weiteren einen Rangkorrelationskoeffizient nach Spearman von  $r = .25$  ( $p \leq .01$ ). Die Variable Handlungswissen ist ebenfalls nicht normalverteilt (Kolmogorov-Smirnov-Test:  $D(192) = .19$ ,  $p \leq .01$ ).

Die Ergebnisse der Evaluation der VSEs werden in Tabelle 4 dargestellt, wobei sich die Studierenden mehrheitlich für eine Fortführung der VSE aussprechen und den Praxisbezug als relativ hoch bewerten. Da die OSCE bereits im Vorfeld erprobt und evaluiert war, soll hier nur in Kürze berichtet werden, dass diese auch im WS18/19 mit einer Gesamtnote von  $Mdn = 2$ ,  $IQR = 1$  als „gut“ bewertet wurde ( $N = 145$ ).

### 4. Diskussion

Das Ziel der erstmaligen Entwicklung und Erprobung einer videobasierten E-Klausur zur Prüfung von Fakten- und Handlungswissen als Bestandteile kommunikativer Kompetenz von Medizinstudierenden im deutschsprachigen Raum konnten wir realisieren. Die Durchführung mit ihrer hohen Komplexität der technischen Abläufe verlief erfreulicherweise nahezu reibungslos. Es ergab sich wie erwartet ein Zusammenhang von Videoklausurergebnis mit der praktischen Prüfung OSCE, der allerdings faktisch gering ausfiel. Eine Betrachtung des Zusammenhangs allein der Items zu Handlungswissen der VSE (nach theoretischer Konzeption) und der OSCE fiel in der Höhe nahezu identisch aus, wobei zu vermuten ist, dass die höhere Bandbreite der Themen in der VSE gegenüber der OSCE dieses Ergebnis erklären kann. Mutmaßlich überdeckt die Themenvielfalt der VSE die Einordnung nach Kompetenzebenen (siehe dazu auch noch Ausführungen weiter unten). Es ist darüber hinaus in Bezug auf die eingangs dargestellte Komplexität des Konstrukts kommunikative Kompetenz und der dargestellten Kompetenzebenen anzunehmen, dass neben Unterschieden in der Themenvielfalt auch noch andere Bereiche bzw. andere Kompetenzebenen (neben Fakten- und Handlungswissen) von kommunikativer Kompetenz in der OSCE erfasst werden. Dies spiegelt sich mutmaßlich ebenfalls in der Höhe der Korrelation wider. Den praktischen Bezug der Prüfung durch den Einsatz von Videos schätzten die Studierenden wie erwartet mehrheitlich höher ein als bei rein schriftlichen Prüfungen und die Arzt-Patient-Gespräche wurden als praxisnah erlebt. Insgesamt legen diese Ergebnisse nahe, dass der angestrebte höhere Praxisbezug im Vergleich zu einer rein textbasierten Prüfung erreicht wurde. Kritisch ist dabei jedoch anzumerken, dass die Studierenden keine Erfahrungen mit einer rein schriftlichen Prüfungsform in diesem Themengebiet hat-



Tabelle 3: Hauptachsenanalysen der VSEs im SS18 und WS18/19, Einfaktorenlösungen

Item	VSE SS18		VSE WS18/19	
	Ladung a	Kommunalität h <sup>2</sup>	Ladung a	Kommunalität h <sup>2</sup>
	Faktor: Patientenorientierung/ Empathie		Faktor: Patientenorientierung/ Empathie	
1	.56	.31		
2				
3			.75	.56
4	.64	.40		
5			.42	.18
6				
7			.44	.19
8				
9			.54	.29
10	.45	.21		
11	.85	.72		
12			.65	.42
13	.66	.44	.50	.25
14				
15				
16	.81	.65	.63	.40
17				
18				
19	.66	.44		
20			.52	.27
21			.61	.37
22				
23			.50	.25
24			.58	.34
25			.76	.58
26	.47	.22	.60	.36
27				
28				
29				
30			.77	.59

Anmerkung. Angegeben sind die Ladungen der Items für die gilt  $h^2 \geq .16$  und Absolutbetrag  $a \geq .40$ .

Tabelle 4: Ergebnisse Evaluation

Item	SS18				WS18/19			
	N	Mdn	IQR	ja	N	Mdn	IQR	ja
Sollte die VSE aus Ihrer Sicht beibehalten werden?	142			78%	153			74%
Gesamtnote der VSE (Schulnotensystem 1-6)	143	2	1		138	3	1	
In der videobasierten Überprüfung kommunikativer Fertigkeiten haben die Videos im Vergleich zu herkömmlichen schriftlichen Prüfungen ohne Videoeinsatz einen höheren Praxisbezug der Inhalte erzeugen können.	143	6	1		138	5	2	
In der videobasierten Überprüfung kommunikativer Fertigkeiten wurde durch die Videos der Bezug zur Praxis des ärztlichen Gesprächs gut hergestellt.	143	6	1		138	5	2	

Anmerkung. Insgesamt N=147 Teilnehmer der Evaluation im SS18 und N=192 im WS18/19. Item 3 und 4 in der Tabelle wurden anhand einer siebenstufigen Likert-Skala von 1 = „stimme gar nicht zu“ bis 7 = „stimme vollkommen zu“ bewertet.

ten und im 2. Semester auch noch wenig praktische Erfahrungen vorweisen können, weswegen ihr Urteil nur eingeschränkt aussagekräftig sein kann. Eine Befragung von Experten wäre zusätzlich zukünftig noch sinnvoll. Zudem muss berücksichtigt werden, dass ein Vergleich mit einer rein schriftlichen Prüfung aus Gründen der Zumutbarkeit hinsichtlich der Belastung für die Studierenden nicht noch zusätzlich realisiert wurde, dieser aber wichtig wäre, um eine höhere Praxisbezogenheit der VSE gegenüber rein schriftlichen Prüfungen zu untersuchen. Die VSE kann des Weiteren in dieser Form noch nicht ausreichend die OSCE wiedergeben.

Hinsichtlich der Testgüte der VSEs nehmen wir darüber hinaus an, dass die Durchführungs- und Auswertungsobjektivität wie erwartet als sehr gut bewertet werden kann,

da die Aufgaben sowie die Gesprächssequenzen in diesem Prüfungsformat standardisiert sind und die Auswertung softwaregesteuert erfolgt. Die Objektivität kann daher mutmaßlich als höher im Vergleich zur OSCE eingeschätzt werden, bei der die Rater-Urteile teilweise eher wenig übereinstimmen [17]. Diese Einschätzung erfolgt allerdings auf Grundlage von Plausibilität und ist nicht durch Daten gedeckt.

Die internen Konsistenzen der beiden VSEs und der OSCE fielen gering aus, wobei entgegen der Erwartung die internen Konsistenzen der VSEs nicht oder kaum höher ausfielen als bei der OSCE. Zu interpretieren ist dieses Ergebnis auf Grundlage der jeweiligen Prüfungsinhalte. Die eingesetzte OSCE bestand aus fünf Stationen mit relativ eng umrissenen Themengebieten. Die VSEs ermöglichten

hingegen durch ihr Design in höherem Maße die gesamte Bandbreite der Lerninhalte, inklusive theoretischer Konzepte, abzudecken (vgl. auch Constructive Alignment [32]). Es zeigte sich allerdings, dass bei Berücksichtigung der gesamten Kursinhalte 30 Items offenbar zu wenig sind, um eine hohe interne Konsistenz zu erzielen. Eine Erhöhung der internen Konsistenz der VSEs wäre somit über eine Erhöhung der Anzahl der Items zu den Themenfeldern oder wie bei der OSCE ebenfalls über eine Einschränkung auf weniger Themenfelder realisierbar. Letztendlich scheint es in Anlehnung an Schecker [33] auch eine prinzipielle Frage zu sein, ob bei der Konstruktion eines Tests der Schwerpunkt auf die Bandbreite oder die Spezifität gelegt wird. Er argumentiert nachvollziehbar, dass es für die Erzielung einer konsistenten Skala leichter sei, Items aus einem einzigen Kontext aufzunehmen, jedoch hinsichtlich der Fachdidaktik bei einem komplexen Themengebiet eine größere Bandbreite wünschenswerter wäre. In Bezug auf zukünftige videobasierte Prüfungen sollte diese Frage demnach erneut gestellt und diskutiert werden. Nach Schecker kann ein geringerer Wert der internen Konsistenz in der Höhe der beiden vorliegenden Werte der VSEs jedoch gut toleriert werden, wenn die fachdidaktische Entscheidung zugunsten der Bandbreite ausfällt.

An diesem Punkt soll auch auf die Ergebnisse der Untersuchung der Faktorenstruktur der VSEs eingegangen werden. Es ergibt sich bei beiden Prüfungen anhand der statistischen Ergebnisse allenfalls eine Tendenz hin zu einem Themenblock in Bezug auf Patientenorientierung und Empathie, wobei viele Items auch nicht zu diesem Faktor zählen und es inhaltlich insgesamt über alle Items gesehen Überschneidungen gibt. Betrachtet man die Kursinhalte, ist dieses Ergebnis allerdings gut nachvollziehbar. Einerseits berücksichtigen mehrere Modelle und Anwendungsfelder des Kursus die Themen Patientenorientierung und Empathie. Andererseits zeigt sich entsprechend der bereits erfolgten Darstellung von ärztlicher kommunikativer Kompetenz als komplex ausgestaltetem Konstrukt, dass unterschiedliches Hintergrundwissen sowie verschiedene Aspekte bei den Items einfließen. Es ist plausibel, dass sich manche Aspekte aufeinander beziehen oder in einer Aufgabenstellung gleichermaßen bedeutsam sind. Fügt man zudem die Ergebnisse zu den internen Konsistenzen mit den Ergebnissen der Hauptachsenanalyse zusammen, bestätigt sich das Bild mehrheitlich eher heterogener Items. Eine zweifaktorielle Lösung mit Faktenwissen und Handlungswissen ergab sich entsprechend der dargelegten Ergebnisse nicht. Es ist dabei anzunehmen, dass der Effekt mehrheitlich heterogener Themenfelder eine Zuordnung hinsichtlich verschiedener Kompetenzebenen überlagert. Die Beobachtung, dass in der VSE im WS18/19 deutlich mehr Items in bedeutsamer Höhe auf dem einen Faktor laden, ist vermutlich dahingehend zu erklären, dass sich die veränderte Schwerpunktsetzung der zweiten Klausur auf Handlungswissen bzw. Aufgaben zur Gesprächsführung und damit der Patientenorientierung/Empathie dadurch widerspiegelt. Es ist auch hinsichtlich dieses Punktes für zukünftige

Klausuren kritisch zu reflektieren, ob eine Schwerpunktsetzung auf diesen Themenbereich gewünscht wird. Insgesamt ist davon auszugehen, dass die VSEs die Lerninhalte des Kursus abbilden und dabei verschiedene Aspekte kommunikativer Kompetenz (wie Empathie ausdrücken, Informierung, Strukturierung, Anwendung von SPIKES-Protokoll u.s.w.) als entsprechend komplexes Konstrukt abdecken.

Die statistischen Kennwerte zeigen des Weiteren auf, dass die VSEs insgesamt eher leicht waren und die Items eine niedrige Trennschärfe aufweisen. Ein Ziel für zukünftige Klausuren dieser Art wird es daher sein, die Schwierigkeit zu erhöhen, um auch eine bessere Differenzierung in der Kompetenzbewertung vornehmen zu können.

Von den Studierenden wurde die neue Prüfungsform gut angenommen. Die Mehrzahl der Studierenden sprach sich für eine Fortsetzung des Einsatzes aus. Dabei muss jedoch angemerkt werden, dass die Gründe hierfür nicht erfasst wurden und es beispielsweise möglich ist, dass sich die Studierenden aufgrund einer geringeren Belastung bei der VSE gegenüber der OSCE für eine Beibehaltung ausgesprochen haben. Dies sollte nochmals überprüft werden.

Während die Entwicklung der VSE zunächst wie erwartet ressourcenintensiv war, steht aufgrund einer großen Menge an produzierten Videos für eine Vielzahl an zukünftigen Klausuren nun Prüfungsmaterial zur Verfügung. Dadurch kann mit einer VSE unserer Einschätzung und Erfahrung nach ressourcenschonender als mit einer OSCE die Leistung der Studierenden überprüft werden (Aussage ist allerdings nicht durch Daten untermauert).

## 5. Schlussfolgerung

Abschließend lässt sich für die Praxis ableiten, dass, wenn es notwendig oder erwünscht ist eine Prüfung auf Basis der direkt beobachtbaren Handlungsebene durchzuführen, die OSCE immer noch den Verhältnissen im realen Arzt-Patient-Gespräch am nächsten kommt. Genügt eine Prüfung mit inhaltlicher Annäherung an die konkrete Praxis mit dennoch anzunehmendem höherem Praxisbezug als bei rein schriftlichen Prüfungen, könnte eine videobasierte Prüfung eine passende Prüfungsform sein, die mittel- und langfristig ressourcenschonender als eine OSCE ist sowie zudem vermutlich eine höhere Objektivität erzielen kann. Allerdings beruhen diese Schlussfolgerungen teilweise auf Plausibilität (wie Ressourcenaufwand und Objektivität) und sind nicht durch Daten untermauert. Die interne Konsistenz ist für beide Prüfungen des Weiteren über die Themenbandbreite und Anzahl der Items bzw. Stationen beeinflussbar, wobei eine Erhöhung der internen Konsistenz bei der VSE leichter umsetzbar erscheint als bei der OSCE. Dies bedeutet auch, dass mit einer VSE leichter mehr Lernziele umfassend abgebildet werden können.

Die neu konzipierte videobasierte E-Prüfung stellt somit je nach Bedarf und gestellten Anforderungen trotz genannter Limitationen insgesamt ein vielversprechendes Instru-

ment im Bereich der Überprüfung bestimmter Aspekte kommunikativer Kompetenzen von Medizinstudierenden dar. Voraussetzungen für den weiteren Einsatz sind allerdings insbesondere eine Verbesserung der Schwierigkeit der Items, eine erneute kritische Reflektion der Frage nach der Höhe der internen Konsistenz und ein noch ausstehender direkter Vergleich mit einer rein schriftlichen Prüfung. Zudem erscheinen weitere Erfahrungen und Untersuchungen an anderen Hochschulen sowie die Entwicklung und Erprobung modifizierter Konzepte vor dem Hintergrund der genannten Limitationen notwendig und erstrebenswert.

## Anmerkung

Bei den dargestellten Hypothesen und Ergebnissen handelt es sich um Inhalte aus der Dissertationsarbeit von Frau Stephanie Ludwig.

## Danksagungen

Wir bedanken uns vielmals beim BMBF für die Förderung des Projekts. Darüber hinaus danken wir herzlich Frau Prof. Wermuth, Herrn Prof. Hardt, Herrn Prof. Beutel, Herrn Dr. Schappert, Herrn Dr. Ditter, Herrn Dr. Seifert, allen Prüfern, Simulationspatienten, wissenschaftlichen Hilfskräften, der Lernklinik mit Herrn Thomas Nowak, dem Medienzentrum, dem ZDV und weiteren Beteiligten für Ihre weitreichende Unterstützung bei der Umsetzung unseres Projekts.

## Förderung

Diese Arbeit wurde durch das Bundesministerium für Bildung und Forschung unter dem Förderkennzeichen (grant number) 16DHL1032 unterstützt.

## Steckbrief

**Name des Standorts:** Universitätsmedizin Mainz

**Studienfach/Berufsgruppe:** Humanmedizin

**Anzahl der Lernenden pro Semester:** ca. 200

**Ist ein longitudinales Kommunikationscurriculum implementiert?** Aktuell läuft eine Pilotimplementierung im Rahmen des Projekts LONGKOM, [<https://www.uni-medizin-mainz.de/lernklinik/startseite/projekte/longkom-kommunikative-kompetenzen-von-aerztinnen-und-aerzten.html>]

**In welchen Semestern werden kommunikative und soziale Kompetenzen unterrichtet?** 2. und 5. Semester, ggf. darüber hinaus, Erfassung derzeit im Rahmen des Projekts LONGKOM

**Welche Unterrichtsformate kommen zum Einsatz?** Rollenspiele, Simulationspatientengespräche, Vorlesung, Blended Learning, Beispielvideos

**In welchen Semestern werden kommunikative und soziale Kompetenzen geprüft (formativ oder bestehensrelevant und/oder benotet)?** 2., 5., 9. Semester

**Welche Prüfungsformate kommen zum Einsatz?** OSCE (bestehensrelevant), videobasierte Single-Choice-Prüfung (bestehensrelevant), videobasierte E-Übung (formativ), Essay-Klausur

**Wer (z.B. Klinik, Institution) ist mit der Entwicklung und Umsetzung betraut?** Klinik und Poliklinik für Psychosomatische Medizin und Psychotherapie, Schwerpunkt Medizinische Psychologie und Medizinische Soziologie, Rudolf Frey Lernklinik Mainz und Praxis Dr. B. Schappert Mainz

## Aktuelle berufliche Rolle der Autor\*innen

- Dipl.-Psych. Stephanie Ludwig: Wissenschaftliche Mitarbeiterin (Klinik und Poliklinik für Psychosomatische Medizin und Psychotherapie, Universitätsmedizin Mainz) im BMBF-geförderten Projekt „digiRole“; zudem tätig als approbierte Psychologische Psychotherapeutin.
- Lina Behling, M.A.: Lina Behling ist wissenschaftliche Mitarbeiterin am Zentrum für Qualitätssicherung und -entwicklung der Johannes Gutenberg-Universität Mainz.
- Univ.-Prof. Dr. Uwe Schmidt: Uwe Schmidt ist Professor für Hochschulforschung am Institut für Soziologie der Johannes Gutenberg-Universität Mainz und Leiter des Zentrums für Qualitätssicherung und -entwicklung. Seine Forschungsschwerpunkte liegen im Bereich der empirischen Hochschulforschung und Evaluationsforschung.
- Dr. rer. physiol. Dipl.-Psych. Sabine Fischbeck, MME: Wissenschaftliche Mitarbeiterin und Unterrichtsbeauftragte für das Fach Medizinische Psychologie und Medizinische Soziologie in Mainz, Master of Medical Education und Psychoonkologin.

## Interessenkonflikt

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

## Literatur

1. Härtl A, Bachmann C, Blum K, Höfer S, Peters T, Preusche I, Raski B, Rüttermann S, Wagner-Menghin M, Wunsch A, Kiessling C; GMA-Ausschuss Kommunikative und Soziale Kompetenzen. Desire and reality—teaching and assessing communicative competencies in undergraduate medical education in german-speaking europe—a survey. *GMS Z Med Ausbild.* 2015;32(5):Doc56. DOI: 10.3205/zma000998
2. Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neill P. Evaluations of situational judgement tests to assess non-academic attributes in selection. *Med Educ.* 2012;46(9):850-868. DOI: 10.1111/j.1365-2923.2012.04336.x

3. Fischbeck S, Mauch M, Leschnik E, Beutel ME, Laubach W. Überprüfung ärztlicher kommunikativer Kompetenz mittels einer OSCE bei Studierenden der Medizin im ersten Studienjahr. *Psychother Psychosom Med Psychol.* 2011;61(11):465-471. DOI: 10.1055/s-0031-1291277
4. Bundesministerium für Gesundheit. Beschluss text zum "Masterplan Medizinstudium 2020". Berlin: Bundesministerium für Gesundheit; 2017. Zugänglich unter/available from: [https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/4\\_Pressemitteilungen/2017/2017\\_1/170331\\_Masterplan\\_Beschlusstext.pdf](https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/4_Pressemitteilungen/2017/2017_1/170331_Masterplan_Beschlusstext.pdf)
5. Association of American Medical Colleges. Report I. Learning Objectives for Medical Student Education. Guidelines for Medical Schools. Washington: Association of American Medical Colleges; 1998. Zugänglich unter/available from: <https://www.aamc.org/system/files/c/2/492708-learningobjectivesformedicalstudenteducation.pdf>
6. UK Foundation Programme Office. The Foundation Programme Curriculum 2016. Birmingham: UK Foundation Programme Office; 2016. Zugänglich unter/available from: [https://www.foundationprogramme.nhs.uk/sites/default/files/2018-07/FP\\_Curriculum\\_2016\\_V2%20\(1\)\\_0.pdf](https://www.foundationprogramme.nhs.uk/sites/default/files/2018-07/FP_Curriculum_2016_V2%20(1)_0.pdf)
7. Aspegren K. Beme guide no. 2: Teaching and learning communication skills in medicine-a review with quality grading of articles. *Med Teach.* 1999;21(6):563-570. DOI: 10.1080/01421599978979
8. Baribeau DA, Mukovozov I, Sabljic T, Eva KW, deLottinville CB. Using an objective structured video exam to identify differential understanding of aspects of communication skills. *Med Teach.* 2012;34(4):e242-250. DOI: 10.3109/0142159X.2012.660213
9. Laidlaw A, Salisbury H, Doherty EM, Wiskin C. National survey of clinical communication assessment in medical education in the united kingdom (UK). *BMC Med Educ.* 2014;14:10. DOI: 10.1186/1472-6920-14-10
10. Chong L, Taylor S, Haywood M, Adelstein BA, Shulruf B. Examiner seniority and experience are associated with bias when scoring communication, but not examination, skills in objective structured clinical examinations in australia. *J Educ Eval Health Prof.* 2018;15:17. DOI: 10.3352/jeehp.2018.15.17
11. Daniels VJ, Harley D. The effect on reliability and sensitivity to level of training of combining analytic and holistic rating scales for assessing communication skills in an internal medicine resident osce. *Patient Educ Couns.* 2017;100(7):1382-1386. DOI: 10.1016/j.pec.2017.02.014
12. Dong T, LaRochelle JS, Durning SJ, Saguil A, Swygert K, Artino AR. Longitudinal effects of medical students' communication skills on future performance. *Mil Med.* 2015;180(4 Suppl):24-30. DOI: 10.7205/MILMED-D-14-00565
13. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9 Suppl):63-67. DOI: 10.1097/00001888-199009000-00045
14. Kopp V, Möltner A, Fischer MR. Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. *GMS Z Med Ausbild.* 2006;23(3):Doc50. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000269.shtml>
15. Schubert S, Ortwein H, Dumitsch A, Schwantes U, Wilhelm O, Kiessling C. A situational judgement test of professional behaviour: Development and validation. *Med Teach.* 2008;30(5):528-533. DOI: 10.1080/01421590801952994
16. Koczwara A, Patterson F, Zibarras L, Kerrin M, Irish B, Wilkinson M. Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Med Educ.* 2012;46(4):399-408. DOI: 10.1111/j.1365-2923.2011.04195.x
17. Chenot JF, Ehrhardt M. Objective structured clinical examination (OSCE) in der medizinischen Ausbildung: Eine Alternative zur Klausur. *Z Allg Med.* 2003;79(9):437-442. DOI: 10.1055/s-2003-43064
18. Schmitz FM, Schnabel KP, Bauer D, Bachmann C, Woermann U, Guttormsen S. The learning effects of different presentations of worked examples on medical students' breaking-bad-news skills: A randomized and blinded field trial. *Patient Educ Couns.* 2018;101(8):1439-1451. DOI: 10.1016/j.pec.2018.02.013
19. Humphris GM, Kaney S. The objective structured video exam for assessment of communication skills. *Med Educ.* 2000;34(11):939-945. DOI: 10.1046/j.1365-2923.2000.00792.x
20. Hulsman RL, Mollema ED, Oort FJ, Hoos AM, de Haes JC. Using standardized video cases for assessment of medical communication skills: Reliability of an objective structured video examination by computer. *Patient Educ Couns.* 2006;60(1):24-31. DOI: 10.1016/j.pec.2004.11.010
21. Fischbeck S, Mauch M, Unterrainer J. Vom Kompetenzwissen zum praktischen Können: Arzt-Patient-Kommunikation prüfen mit Key-Feature-Klausuren und OSCE. In: Berth H, editor. *In balance. Abstracts zur Jahrestagung der Deutschen Gesellschaft für Medizinische Psychologie 2013.* Lengerich: Pabst-Publishers; 2013. p.126-127.
22. Kurtz S, Silverman J, Draper J. *Teaching and learning communication skills in medicine.* 2nd ed. Oxford: Radcliffe; 2006.
23. Back AL, Arnold RM, Baile WF, Fryer-Edwards KA, Alexander SC, Barley GE, Gooley TA, Tulsy JA. Efficacy of communication skills training for giving bad news and discussing transitions to palliative care. *Arch Intern Med.* 2007;167(5):453-460. DOI: 10.1001/archinte.167.5.453
24. Kanfer FH, Reinecker H, Schmelzer D. *Selbstmanagement-Therapie.* 5. Aufl. Berlin, Heidelberg: Springer; 2012. DOI: 10.1007/978-3-642-19366-8
25. Hamann J, Loh A, Kasper J, Neuner B, Spies C, Kissling W, Härter M, Heesen C. Partizipative Entscheidungsfindung. Implikationen des "shared decision making" für Psychiatrie und Neurologie. *Nervenarzt.* 2006;77(9):1071-1078. DOI: 10.1007/s00115-005-1950-5
26. Baile WF, Buckman R, Lenzi R, Glober G, Beale EA, Kudelka AP. Spikes-a six-step protocol for delivering bad news: Application to the patient with cancer. *Oncologist.* 2000;5(4):302-311. DOI: 10.1634/theoncologist.5-4-302
27. Hoppe-Seyler T, Gartmeier M, Möller G, Bauer J, Wiesbeck A, Karsten G. Entwicklung von Lehrfilmen zur Gesprächsführung zwischen Realitätsnähe und systematischer didaktischer Gestaltung. *ZFHE.* 2014;9(3):127-135. DOI: 10.3217/zfhe-9-03/14
28. Edwards AW. The Measure of Association in a 2 × 2 Table. *J Royal Stat Soc.* 1963;126(1):109-114. DOI: 10.2307/2982448
29. Leonhart R. *Lehrbuch Statistik.* 2. Aufl. Bern: Huber; 2009.
30. Wermuth N. How can graphical Markov models aid causal inference? In: *DAGStat 2019. Statistics under one umbrella. Conference Guide.* München, 18.-20. März 2019. Göttingen: Deutsche Arbeitsgemeinschaft Statistik; 2019.
31. Rost DH, Schermer FJ. Strategien der Prüfungsangstverarbeitung. *Z Diff D P.* 1986;7(3):127-139.
32. Biggs J. Enhancing teaching through constructive alignment. *High Educ.* 1996;32(3):347-364. DOI: 10.1007/BF00138871



33. Schecker H. Überprüfung der Konsistenz von Itemgruppen mit Cronbachs alpha. In: Krüger D, Parchmann I, Schecker H, editors. Methoden in der naturwissenschaftsdidaktischen Forschung. Berlin: Springer Spektrum; 2014. chapter online-Zusatzmaterial. DOI: 10.1007/978-3-642-37827-0\_1

**Korrespondenzadresse:**

Dipl.-Psych. Stephanie Ludwig  
Universitätsmedizin der Johannes Gutenberg-Universität  
Mainz, Klinik und Poliklinik für Psychosomatische Medizin  
und Psychotherapie, Schwerpunkt Medizinische  
Psychologie und Medizinische Soziologie, Saarstr. 21  
(Campus Universität), 55099 Mainz, Deutschland  
stephsch@uni-mainz.de

**Bitte zitieren als**

Ludwig S, Behling L, Schmidt U, Fischbeck S. Development and testing of a summative video-based e-examination in relation to an OSCE for measuring communication-related factual and procedural knowledge of medical students. *GMS J Med Educ.* 2021;38(3):Doc70. DOI: 10.3205/zma001466, URN: urn:nbn:de:0183-zma0014667

**Artikel online frei zugänglich unter**

<https://www.egms.de/en/journals/zma/2021-38/zma001466.shtml>

**Eingereicht:** 29.01.2020

**Überarbeitet:** 25.10.2020

**Angenommen:** 20.11.2020

**Veröffentlicht:** 15.03.2021

**Copyright**

©2021 Ludwig 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/>.