

Development, testing and generalizability of a standardized evaluation form for the assessment of patient-directed reports in the new final medical licensing examination in Germany

Abstract

Background: As doctors often fail to explain diagnoses and therapies to patients in an understandable and appropriate way, the improvement of doctor-patient communication is essential. The current medical training and examinations are focused on verbal rather than on written communication. Following the premise of “assessment drives learning”, the final medical licensing examination in Germany has been further developed by the German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy (IMPP). As part of the discharge management the candidates have to prepare a report for the patient that is understandable and provides them with all important information about their stay in hospital.

Aim: A standardized evaluation form for formative and summative feedback has been developed and tested with regard to applicability and the assurance of test quality criteria, especially the reliability to assess the written communication skills of the students.

Methodology: In an expert consensus procedure, a draft for a standardized evaluation form was developed. This form was revised after an initial trial run on patient-directed reports written by students in their last year of medical studies. Afterwards twenty-one patient-directed reports were evaluated by fourteen different examiners. The reliability was tested by calculating the generalizability-coefficient and by analysing the inter-rater reliability.

Results: The first test on the evaluation of the patient-directed reports indicated the practicability of the application and the usefulness of the evaluation form as an instrument for assessing the written communication skills of students. The analyses of the inter-rater reliability showed that the degree of agreement in the evaluations was partly different between two groups of examiners. The calculated G-coefficient indicates a high reliability. The content validity of the evaluation form was given through the comprehensive medical expertise in the development process.

Conclusion: Assessing written patient-directed communication is a benefit of the newly developed last part of the medical licensing examination in Germany. Continuous formative assessment and feedback based on the evaluation form is intended to improve the written communication skills of future doctors. Furthermore, a better understanding of their diagnosis and treatment as well as a trusting relationship with their doctor may empower patients in the medical decision process and lead to fewer dismissal errors in the future. For consistent use of the evaluation form a standardized training of examiners should be implemented.

Keywords: communication, education, patient participation

Lena Selgert¹
Bernd Bender¹
Barbara Hinding¹
Aline Federmann¹
André L. Mihaljevic²
Rebekka Post³
Ansgar Jonietz³
John Norcini⁴
Ara Tekian⁵
Jana Jünger¹

1 Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP), Mainz, Germany

2 Universitätsklinikum Heidelberg, Klinik für Allgemein-, Viszeral- und Transplantationschirurgie, Heidelberg, Germany

3 "Was hab' ich?" gGmbH, Dresden, Germany

4 SUNY Upstate Medical University, Department of Psychiatry, New York, USA

5 University of Illinois at Chicago, College of Medicine, Illinois, USA

1. Introduction

Following the “patient’s rights law” by the German civil law code every patient has the right to be fully informed [1]. Nevertheless, several studies show that doctors often fail to explain diagnoses and therapies to patients in an understandable and appropriate way. Therefore an improvement of doctor-patient communication is essential: 22% of patients receive incomprehensible answers to their questions and in 29% incomprehensible explanations of examination results from their doctors. As a result, 39% of patients feel left alone with their worries and fears [2].

Incorrect communication by medical professionals caused up to 33% of dismissal errors shown in a study between October 2012 and September 2013 [3]. Poor communication during discharge lead to medication errors, poor wound care, inadequate nutrition, rehospitalization, life-threatening situations, avoidable and unnecessary medical services and procedures as well as additional work for nursing services and increased costs for the health system. Providing sufficient and written information for patients is essential to guarantee their adherence to therapy and to implement preventive measures [2], [4], [5], [6], [7]. Especially communicating without using medical terminology is emphasized as a meaningful strategy to empower patients in their decision making process [8].

To improve the doctor-patient-communication of future doctors, Jünger et al. developed a longitudinal communication curriculum [9], [10]. Based on this curriculum the essential learning objectives of doctor-patient-communication can be integrated into medical training and assessment [11].

As assessment drives learning [12], the final part of the medical licensing examination in Germany was further developed by the German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy (IMPP) to meet the needs of patients in order to optimally prepare medical students for their first day at work. Common assessments of communication skills e.g. objective structured clinical examinations, often used simulated patients and focused on verbal communication skills [13]. To increase the authenticity the new workplace-based examination demands real patients: on a surgical or internal medicine ward as well as in outpatient care (see figure 1). As part of the improved discharge management candidates will prepare an evidence-based patient report for the post-discharge attending physician and a report for the patients themselves that is easily understandable and provides them with all important information [14].

Use of simple language is one of the most common strategies used by doctors, nurses and pharmacists in order to improve communication with their patients [15]. Earl et al. showed the impact of a health literacy module on the improvement of students' written patient education materials in the areas of readability, message content, computational power, statistics and concepts of patient

activity. On the other hand, the simplification of medical language remained difficult [16]. Summarizing, there is already a number of international studies that have dealt with doctor-patient-communication in general and written communication in particular [17], [18], [19], [20], [21]. So far, these studies have dealt primarily with the relationship between patients' health literacy and communication, the legibility and comprehensibility of written patient information materials, as well as the benefits of written communication strategies. A larger randomized, controlled study based on an understandable patient report was already conducted with 417 patients by the initiative “Was hab' ich?” gGmbH. The physicians at “Was hab' ich?” provided patients with an easily understandable patient report after discharge from hospital. Their study showed significant effects of the patient reports on the patients' understanding of examination results, medication indications and prescriptions [22]. The language used in the reports was characterized by simple words, short, complete and simple sentences as well as positive language and the avoidance of medical terms. Relevant background information was provided and the text had a logical structure [23], [24]. On the basis of this collection of criteria for the preparation of a report in patient-directed language, “Was hab' ich?” designed a template for the preparation of such reports by medical students. So far, there is no evaluation instrument for these patient-directed reports.

The aim of this study was to develop and to test a standardized evaluation form for formative and summative assessment of reports which covered the important aspects of patient-directed writing. Based on real rather than simulated patients and situations, this form had to be used individually and to cover all different kinds of settings and diseases.

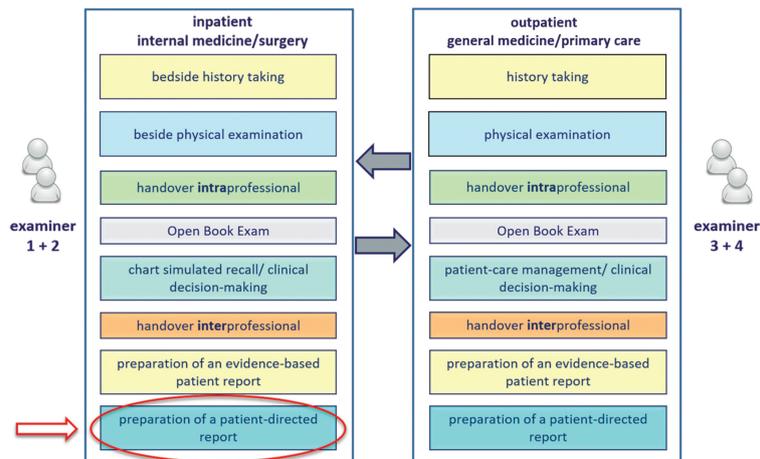
This evaluation form was based on the collection of criteria by “Was hab' ich?” gGmbH, literature analysis and expert opinions. The applicability of the form and the test quality criteria especially the reliability were tested.

2. Methodology

2.1. Development of a standardized evaluation form for patient-directed reports

In order to assess the quality of written communication in patient-directed reports by medical students, a first draft of an evaluation form was drawn up based on a literature analysis and on a collection of important criteria for writing patient-directed by “Was hab' ich?” gGmbH and the IMPP [23], [24]. In August 2018, a group of 27 medical experts from eight German faculties consisting of specialists in general medicine, internal medicine, anaesthesia, psychiatry, surgery, psychosomatic medicine and psychotherapy, psychologists, who were all participants in the German “Master of Medical Education” (MME) study programme, and of five students prepared a precise evaluation form in a consensus procedure: The

New final part of the medical licensing examination in Germany



Under the supervision of two examiners, relevant medical skills are assessed in eight steps: starting with a complete medical history taking and physical examination of a patient who was previously unknown. The collected information are passed on to the examiners in a structured manner. Subsequently, the students answer a specific clinical question related to the patient according to the PICO scheme (Patient population – Intervention – Comparison – Outcome) with the help of scientific literature research (Open Book Exam) and fill out a standardized patient curve. This curve is presented to the investigators and serves to reflect clinical decisions. Afterwards, the patient is handed over to a responsible person from the interprofessional treatment team to demonstrate interprofessional competencies. Finally, the students prepare an evidence-based patient report as well as a patient-directed report, which provides the patients with all important information about their stay in hospital [14].

Figure 1: Workplace-based examination on real patients on a surgical or internal medicine ward and in the outpatient area

first draft was revised first in a small group of seven experts and afterwards discussed in the whole group until a consensus was found.

In October 2018, the evaluation form was tested on a total of ten students at the Department of Surgery of the University Hospital of Heidelberg [25]. These students were in their 4th till 6th year of medical studies. Each student wrote a patient-directed report by filling out a standardized template. This report was assessed by eleven physicians and three students who had been involved in the development of the evaluation form in August 2018. Based on these experiences the evaluation form was revised by the participants of the initiative “Was hab’ ich?” gGmbH and the IMPP.

2.2. Test and revision of the evaluation form

In January and February 2019, the revised evaluation form was tested extensively on twenty-one patient-directed reports written by students in their last (6th) year of medical studies (PJ) at the interprofessional training ward (HIPSTA) of the University Hospital of Heidelberg. At this HIPSTA, medical students and trainees from different health professions treat patients together under the supervision of medical and nursing facilitators for four weeks [26]. One part of the practical training is writing patient-directed reports in addition to the conventional discharge report to the attending physician. The PJ-students receive a structured training in patient-directed writing by the initiative “Was hab’ ich?” gGmbH. This training includes elements like the reliable recognition and avoidance of medical terms, the explanation of background information and simply structured writing.

These twenty-one patient-directed reports were evaluated by two groups of examiners, each consisting of two physicians from the initiative “Was hab’ ich?” gGmbH. The first group used the developed evaluation form with the precise sub-items. The second group that evaluated the reports focused on the three main criteria without knowledge of the more precise sub-items [27].

2.3. Evaluation study

After revision of the evaluation form based on the first testing at HIPSTA, the twenty-one patient-directed reports were evaluated by a total of fourteen examiners. These examiners consisted of members from the initiative “Was hab’ ich?” gGmbH who were experienced in patient-directed writing and of physicians from the IMPP as well as general practitioners, as these are the physicians who usually receive discharge reports. The corresponding discharge report for the attending post-discharge physicians was available for comparison. The sample used for the statistical analyses included all existing valuations of all examiners on the twenty-one reports. Since not all examiners evaluated all reports, a sample of $n=205$ valuations resulted. The sample for the analyses on inter-rater reliability consisted of the evaluations of nine examiners, who had fully evaluated eleven of the twenty-one reports. As no personal data were evaluated, no approval of the ethics committee was needed.

2.4. Test quality criteria and statistical methods

Statistically, descriptive statistics in the form of mean values and standard deviations were first analysed sep-

arately for each of the three categories. In addition, bivariate correlations between the categories were calculated. Therefore, Pearson's correlation coefficient was used [28].

Reliability is given when the respective instrument does not produce any measurement errors. It can be assumed if there is a high degree of intercorrelation between the individual parts of a measuring instrument. The reliability was analysed by the calculation of the generalizability coefficient [29]. Numerous measurements by the same examiners lead to an overestimation of reliability due to exercise effects when using Cronbach's alpha [30]. In order to correct that, variance components were calculated for the factors examiner, report and background of the examiner ("Was hab' ich?" vs. IMPP and general practitioners) regarding the reached score given in all valuations based on the generalizability theory [29], [31]. This helped to identify possible sources of measurement errors in the evaluations of the reports [32]. The relative error variance is determined based on the calculated variance components. With this relative error variance the G-coefficient can be computed. The G-coefficient estimates if the results can be transferred to the study population or if the interaction effect between the facets and the participants make the results singular to the study sample. A G-coefficient of 1 indicates that the available data and results can be perfectly generalized to all evaluations outside the study. A high value for this coefficient thus indicates high reliability [29].

The evaluation agreement in the form of inter-rater reliability [33] was calculated on the basis of a sample of nine examiners, who fully evaluated eleven of the twenty-one reports. The other examiners evaluated only some of these reports. The inter-rater reliability was handled separately for each of the three categories across the eleven reports. The analysis of inter-rater reliability was based on the coefficient Kendall W [34]. The evaluation scale of the developed form consisted of a rating scale ascending from 0 to 5 points. As at least an ordinal scale level could be assumed, Kendall W seemed to be the appropriate coefficient, unlike Fleiss Kappa for example, which requires categorical data [34]. The aim was to determine whether or not the examiners were consistent in their valuations within one evaluation category.

Furthermore, the content validity of the evaluation form was given through the comprehensive medical expertise in the development process. This means that the developed categories of the evaluation form represent exactly the content that is intended [35]. To avoid biases the examiners did not know the students who had prepared the patient-directed reports personally.

3. Results

3.1. Development and first test of the evaluation form

This first draft contained the categories "Selection of content", "Medical correctness", "Report structure and syntax", "Linguistic design" and "Grammar". Each category could be awarded with zero to five points. Each of the five categories was weighted with ten to thirty percent. Resulting from the first test, the main evaluation criteria were summarized into three: "Selection of content and medical correctness", "Transfer of medical language into lay language" and "Easily understandable language". The selection of content was presented more clearly and differentiated. The three central evaluation categories were provided with specific sub-items to explain the rating categories in detail. Each category could be awarded with zero to five points. The category "Selection of content and medical correctness" was given a slightly higher weighting of forty percent. The other two categories were given a weighting of thirty percent each.

Following the first testing at HIPSTA, the previous version of the evaluation form was revised and specified. The second category was renamed "Lay language and background information". To reduce the number of sub-items especially those of the category "selection of content and technical correctness" have been summarized from eight to five.

3.2. Tested version of the evaluation form

A standardized evaluation form for patient-directed reports has been created successfully with three main evaluation categories. The percentage weighting remained forty percent for "Selection of content and medical correctness" and thirty percent each for the categories "Lay language and background information" and "Easily understandable language".

The sub-items served to specify the content of these evaluation categories (see figure 2).

3.3. Evaluation study

3.3.1. Descriptive statistics and correlations

Based on the sample of 205 individual valuations by the examiners, mean values, standard deviations and paired correlations between the three categories were calculated.

First, an overview of mean values and standard deviations for each of the three categories of the evaluation form can be provided (see table 1).

The mean values and standard deviations of the three categories were almost identical. This clearly showed that, on average, the same number of points was awarded across the evaluations by all fourteen examiners for the twenty-one reports. Thus, no category was rated better or worse on average.



		Points
Selection of content and medical correctness (40%)		0-5
<ul style="list-style-type: none"> - selects the information relevant for the patient - reproduces the information professionally correct 		
The reason for your stay in hospital	<ul style="list-style-type: none"> - Explains the reason for admission in a few sentences including complaints, previous diagnostics and therapy 	
Your medical history and your previous diseases	<ul style="list-style-type: none"> - Describes important results of the patient medical history and the physical examination - Explains important secondary diagnoses 	
The course of your stay in hospital	<ul style="list-style-type: none"> - Explains conclusively the course of the disease with diagnostics and therapy carried out - Explains any complications 	
Your medication	<ul style="list-style-type: none"> - Explains the indication/desired effect in each case - Explains the regimen of use 	
The next steps	<ul style="list-style-type: none"> - Provides information on further planned diagnostics and therapy including appointments, rehabilitation, etc.. - Gives behavioural recommendations (hygiene, wound care, nutrition) and explains concrete procedures for complications. 	
Lay language and background information (30%)		0-5
<ul style="list-style-type: none"> - Finds and uses good equivalents for technical terms throughout the text - Explains difficult words and provides appropriate background information - Explains relevant anatomical and physiological contexts - Uses only common abbreviations 		
Easily understandable language (30%)		0-5
<ul style="list-style-type: none"> - Addresses the patient directly - Formulates whole, short and simple sentences - Uses simple words from everyday language - Avoids elevated language style (e.g. nouns, subjunctive, passive, imperfect) - Writes orthographically and grammatically correct 		
Overall rating		/100%

Figure 2: Tested evaluation form for assessment of patient-directed reports

Table 1: Descriptive statistics of the three evaluation categories

Category	Mean (Standard deviation)
Selection of content and technical correctness	3.8 (0.99)
Lay language and background information	3.8 (0.97)
Easily understandable language	3.7 (0.95)

Descriptive statistics (mean and standard deviation) for each of the three categories based on the sample of n=205 valuations in total.

Since the results for the descriptive statistics already suggested a possible correlation, this was checked using bivariate correlations between the three categories (see table 2).

The results of the pairwise correlations showed medium and high positive correlations between the three categories, which were all highly significant (p-value <0.001). The strongest correlation was found between the categories “Lay language and background information” and “Easily understandable language” (r=0.61). In contrast, the correlations between “Selection of content and technical correctness” and “Lay language and background information” (r=0.45) and between “Selection of content and technical correctness” and “Easily understandable language” (r=0.31) were medium strong.

3.3.2. Reliability

The inter-rater reliability was handled separately for each of the three categories across the eleven reports. This was done for the examiners from “Was hab’ ich?” and for those from the IMPP and general practitioners, in order to compare the degree of agreement of these two groups (see table 3).

The degree of agreement in the first category was substantially higher for the examiners of “Was hab’ ich?” compared to the examiners from the IMPP and general practitioners. Regarding the two other categories the degree of agreement was moderate in both groups.

The **calculated G-coefficient was 0.72 based on all 205 valuations**. This rather high value is an indicator that the evaluation results of the patient-directed reports are not

Table 2: Bivariate correlations between the three categories

Category	Pearson's r
Selection of content and technical correctness - Lay language and background information	0.45***
Selection of content and technical correctness - Easily understandable language	0.31***
Lay language and background information - Easily understandable language	0.61***

Bivariate correlations calculated by Pearson's correlation coefficient between the three categories based on the sample of n=205 valuations in total. The classification of the values for Pearson's correlation coefficient according to Kühnel & Krebs [36] are: 0.00<r<0.05=no correlation; 0.05<r<0.20=low correlation; 0.20<r<0.50=medium correlation; 0.50<r<0.70=high correlation; r>0.70=very high correlation. The significance level used was 0.05. Legend: n.s.=no significance, *p-value<0.05; **p-value<0.01; ***p-value<0.001.

Category	Examiners of "Was hab ich?" (n = 5)	Examiners from IMPP and general practitioners (n = 4)
	Kendall W	Kendall W
selection of content and technical correctness	0.82	0.35
lay language and background information	0.51	0.58
easily understandable language	0.64	0.65

Inter-rater-reliability of the 11 patient-directed reports. Kendall concordance coefficient: 0=No agreement, 1=perfect agreement. The classification of the values for Kendall W according to R. C. Schmidt [37] are: 0.1 - very weak agreement; 0.3 - weak agreement; 0.5 - moderate agreement; 0.7 - strong agreement; 0.9 - exceptionally strong agreement.

Table 3: Inter-rater reliability separately for each category across the eleven reports

limited to the sample of the study, but can be transferred to evaluations outside the study.

3.3.3. Revision of the evaluation form based on the evaluation

Based on these results the evaluation form was slightly revised: The category "Lay language and background information" was specified to "Provision of background information and patient-understandable use of technical terms". The category "Easily understandable language" was specified to "Patient-understandable language style, readability and everyday speech" to enable better differentiation of these two categories.

The explanation of the item "your medication" was complemented with "Explains the intake schedule, gives intake instructions." and "Indicates relevant interactions and/or adverse effects." as well as the explanation of the behavioural recommendations of "The next steps" with "hygiene, wound care, nutrition, exercise, drinking quantity, nicotine consumption".

4. Discussion

A multi-stage conception and revision process was successfully used to create a standardized evaluation form for assessment of patient-directed reports. This is a noticeable improvement of the students training in the field of doctor-patient-communication.

It has been shown, that medical students who have undergone written communication training and regularly translate findings use better explanations than untrained students when talking to standardized patients in a simulated physician-patient contact [36]. This corresponds to the self-awareness of the students working at "Was hab' ich" GmbH: they are united in their opinion that written translations of doctors' reports to patient-reports

improve their ability to communicate in a way that can be understood better by patients [37].

In this study students had to prepare patient-directed reports in their last year of medical training. The test of a newly developed evaluation form used on twenty-one of these patient-directed reports showed the practicability as well as the usefulness of the instrument for assessing these written communication skill of the students. The individual evaluation categories represent the most important steps of writing a patient-directed report as the evaluation form has been developed by different medical experts. The implementation of specific sub-items supports the examiners in interpreting the categories.

Medium or high correlations between the three categories could be observed. Above all, the high correlation between lay language and patient-understandable language is an indicator of a good reliability of the evaluation form. In contrast, the medium-strong correlations of the content selection with lay language as well as with patient-understandable language are an indication that the evaluations of content and language can be independent of each other. Those students who, in the opinion of the examiners, used appropriate lay language in the patient reports and communicated the background information well were also able to write in language that was understandable to the patient. In contrast, an appropriate selection of content and technical correctness was not necessarily dependent on the use of lay language or easily understandable language.

The analyses of inter-rater reliability showed that the degree of agreement in the evaluations for all three categories was partly different between the two groups of examiners. The examiners of the initiative "Was hab' ich?" had a higher level of agreement in the category "selection of content and medical correctness" than the examiners of the IMPP and the general practitioners. This fact is an indication for different starting conditions of the examiners. This can be attributed to their different disciplinary

backgrounds and especially their different previous experiences with patient-understandable language. This finding highlights the need for an uniform training of examiners on standards for writing a patient-directed report before using the evaluation form in the national licensing examination.

The calculated G-coefficient showed that the data and results of the study can be applied to evaluations outside the study and were therefore generalizable, which showed the reliability of the evaluation form.

As the content validity of the developed evaluation form was given, the test quality criteria were mostly fulfilled. Thus, the evaluation form was a useful instrument to rate patient-directed reports with central evaluation categories. Some limitations include that the results were only obtained from reports written at one interprofessional training ward in Heidelberg. In the context of further research, it would be interesting to evaluate and validate the performance of assessment based on the evaluation of a larger sample of examiners and reports written on a conventional ward or in the outpatient area of different faculties. Examiner trainings on assessing patient-directed reports using the developed evaluation form should be mandatory. This training is intended to ensure a uniform assessment standard to avoid different interpretations and weightings of sub-items and to contribute to a higher assessment agreement. The inter-rater reliability of examiners who have been trained should be analysed.

Continuous formative assessment and feedback based on the evaluation form would improve the medical training for PJ-students. As the patients are the receivers of these reports, the comparison of the assessment of the patient-directed reports by the patients themselves to the assessment by medical experts would be of great interest. Combining these two feedbacks can ensure that the patient-directed reports will deliver the important information to avoid dismissal errors in the future.

The improvement of the verbal communication skills based on this improvement of medical training should be investigated in further studies.

5. Conclusion

Assessing written patient-directed communication is a benefit of the new developed last part of the medical licensing examination in Germany. Continuous formative assessment and feedback based on the evaluation form is intended to improve the written communication skills of future doctors. Furthermore, a better understanding of their diagnosis and treatment as well as a trusting relationship with their doctor may empower patients in the medical decision process and lead to fewer dismissal errors in the future. To reach this goal, clear instructions and training for writing a patient-directed report must be part of the medical curriculum. For consistent use of the evaluation form a standardized training of examiners should be implemented.

List of abbreviations

- G-coefficient = Generalizability coefficient
- HIPSTA = Interprofessional training ward of the University Hospital of Heidelberg
- IMPP = German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy
- MME = Master of Medical Education
- PJ = Elective clerkships in the final (6th) year
- r = Pearson's correlation coefficient r

Funding

This project was funded by the Bertelsmann Stiftung (duration: 1.10.2017 – 30.06.2021).

Current professional roles of the authors

Dr. med. Lena Selgert

- Physician
- Research assistant at the German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy (IMPP)

Bernd Bender

- Graduate sociologist
- Research assistant at the German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy (IMPP)

Dr. phil. Barbara Hinding

- Graduate psychologist
- Research assistant at the German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy (IMPP)
- Research areas: medical conversation and interprofessional communication in teaching and assessment, implementation of communication curricula in medical education and advanced training.

Aline Federmann, M.A.

- Graduate sociologist
- Research assistant at the German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy (IMPP)

Prof. Dr. med. André L. Mihaljevic

- Physician and medical study program coordination (Department of General, Visceral and Transplant Surgery, Heidelberg University Hospital)
- Medical director of the interprofessional training ward in Heidelberg - (HIPSTA)
- Clinical scientist (deputy spokesman of the surgical study network CHIR-Net)

Rebekka Post

- physician working at the initiative "Was hab' ich?" gGmbH

Ansgar Jonietz

- co-founder and CEO of the initiative "Was hab' ich?" gGmbH
- computer scientist

John J. Norcini, Ph.D.

- Research Professor in the Department of Psychiatry at SUNY Upstate Medical University
- President Emeritus of FAIMER

Ara Tekian, Ph.D., MHPE

- professor, Department of Medical Education, and associate dean, International Education, University of Illinois at Chicago College of Medicine, Illinois, USA.

Prof. Dr. med. Jana Jünger, MME (Bern)

- Director of the German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy (IMPP)
- Development of the post-graduation study program Master of Medical Education (MME), Germany
- Member of the MME-study program management and lecturer for the modules Assessment, Education Research and Evaluation
- Management of various programs on the implementation of communication curricula in medical training and the development of new examination formats for assessing communicative skills.

Competing interests

The authors declare that they have no competing interests.

References

1. Bundesministerium für Gesundheit. Patientenrechte. Berlin: Bundesministerium für Gesundheit; 2019. Zugänglich unter/available from: <https://www.bundesgesundheitsministerium.de/themen/praevention/patientenrechte/patientenrechte.html>
2. Stahl K, Nadj-Kittler M. Picker report 2016. Vertrauen braucht gute Verständigung. Erfolgreiche Kommunikation mit Kindern, Eltern und erwachsenen Patienten. Hamburg: Picker Institut Deutschland gGmbH; 2016.
3. NHS England. Review of National Reporting and Learning System (NRLS) incident data relating to discharge from acute and mental health trusts - August 2014. London: NHS England; 2014.
4. Williams H, Edwards A, Hibbert P, Rees P, Prosser Evans H, Panesar S, Carter B, Parry G, Makeham M, Jonas A, Avery A, Sheikh A, Donaldson L, Carson-Stevens A. Harms from discharge to primary care: mixed methods analysis of incident reports. *Br J Gen Pract.* 2015;65(641):829-837. DOI: 10.3399/bjgp15X687877
5. Hesselink G, Zegers M, Vernooij-Dassen M, Barach P, Kalkman C, Flink M, Öhlen G, Olsson M, Bergenbrant S, Orrego C, Suñol R, Toccafondi G, Venneri F, Dudzik-Urbaniak E, Kutryba B, Schoonhoven L, Wollersheim H; European HANDOVER Research Collaborative. Improving patient discharge and reducing hospital readmissions by using Intervention mapping. *BMC Health Serv Res.* 2014;14:389. DOI: 10.1186/1472-6963-14-389
6. Pinelli V, Papp KK, Gonzalo JD. Interprofessional communication patterns during patient discharges: A social network analysis *J. Gen Intern Med.* 2015;30(9):1299-306. DOI: 10.1007/s11606-015-3415-2
7. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in Communication and Information Transfer between Hospital-Based and Primary Care Physicians: Implications for Patient Safety and Continuity of Care. *JAMA.* 2007;297(8):831-841. DOI: 10.1001/jama.297.8.831
8. Noordman J, van Vliet L, Kaunang M, van den Muijsenbergh M, Boland G, van Dulmen S. Towards appropriate information provision for and decision-making with patients with limited health literacy in hospital-based palliative care in Western countries: a scoping review into available communication strategies and tools for healthcare providers. *BMC Palliat Care.* 2019;18(1):37. DOI: 10.1186/s12904-019-0421-x
9. Jünger J, Mutschler A, Kröll K, Weiss C, Fellmer-Drüg E, Köllner V, Ringel N. Ärztliche Gesprächsführung in der medizinischen Aus- und Weiterbildung: Das nationale longitudinale Mustercurriculum Kommunikation. *Med Welt.* 2015;66: 189-192.
10. Sator M, Jünger J. From Stand-Alone Solution to Longitudinal Communication Curriculum -Development and Implementation at the Faculty of Medicine in Heidelberg. *Psychother Psych Med.* 2015;65(05):191-198. DOI: 10.1055/s-0034-1398613
11. Jünger J, Weiss C, Fellmer-Drüg E, Semrau J. Verbesserung der kommunikativen Kompetenzen im Arztberuf am Beispiel der Onkologie: Ein Projekt des Nationalen Krebsplans. *Forum.* 2016;31:473-478. DOI: 10.1007/s12312-016-0162-1
12. van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205-214. DOI: 10.3109/0142159X.2012.652239
13. Nikendei C, Jünger J. OSCE - hands on instructions for the implementation of an objective structured clinical examination. *GMS Z Med Ausbild.* 2006;23(3):Doc47. Zugänglich unter/available from: <https://www.egms.de/static/de/journals/zma/2006-23/zma000266.shtml>
14. Jünger J. Kompetenzorientiert prüfen im Staatsexamen Medizin. *Bundesgesundheitsbl.* 2018;61:171-177. DOI: 10.1007/s00103-017-2668-9
15. Schwartzberg JG, Cowett A, VanGeest J, Wolf MS. Communication techniques for patients with low health literacy: a survey of physicians, nurses, and pharmacists. *Am J Health Behav.* 2007;Suppl 1:96-104. DOI: 10.5555/ajhb.2007.31.supp.S96
16. Earl GL, Harris EM, Dave M, Estriplet-Jiang, J. Implementing a health literacy module fostering patient-centered written communication in a cardiovascular prevention elective course. *Curr Pharm Teach Learn.* 2019;11(7):702-709. DOI: 10.1016/j.cptl.2019.03.008
17. Lopez Ramos C, Williams JE, Bababekov YJ, Chang DC, Carter BS, Jones PS. Assessing the Understandability and Actionability of Online Neurosurgical Patient Education Materials. *World Neurosurg.* 2019;130:588-597. DOI: 10.1016/j.wneu.2019.06.166
18. Roberts HJ, Zhang D, Earp BE, Blazar P, Dyer GSM. Patient self-reported utility of hand surgery online patient education materials. *Musculoskeletal Care.* 2018;16(4):458-462. DOI: 10.1002/msc.1360

19. Davis TC, Fredrickson DD, Arnold C, Murphy PW, Herbst M, Bocchini JA. A polio immunization pamphlet with increased appeal and simplified language does not improve comprehension to an acceptable level. *Patient Educ Couns*. 1998;33(1):25-37. DOI: 10.1016/s0738-3991(97)00053-0
20. Rubin DT, Ulitsky A, Poston J, Day R, Huo D. What is the most effective way to communicate results after endoscopy? *Gastrointest Endosc*. 2007;66(1):108-112. DOI: 10.1016/j.gie.2006.12.056
21. Schumaier AP, Kakazu R, Minoughan CE, Grawe BM. Readability assessment of American Shoulder and Elbow Surgeons patient brochures with suggestions for improvement. *JSES Open Access*. 2018;2(2):150-154. DOI: 10.1016/j.jses.2018.02.003
22. "Was hab' ich?" gGmbH. Patientenbriefe wirken. Ergebnisbericht zum Projekt "Mehr Gesundheitskompetenz durch Patientenbriefe". Hamburg: "Was hab' ich?" gGmbH; 2019.
23. Netzwerk Leichte Sprache. Die Regeln für Leichte Sprache. Münster: Netzwerk Leichte Sprache; 2017. Zugänglich unter/available from: https://www.leichte-sprache.org/wp-content/uploads/2017/11/Regeln_Leichte_Sprache.pdf
24. Bredel U, Maaß C. Duden Leichte Sprache. Theoretische Grundlagen, Orientierung für die Praxis. Berlin: Dudenverlag; 2016.
25. Selgert L, Samigullin A, Lux R, Gornostayeva M, Hinding B, Schlasius-Ratter U, Hendelmeier M, Mihaljevic AL, Wienand S, Schneidewind S, Bintaro P, Jonietz A, Jünger J. Weiterentwicklung des medizinischen Staatsexamens in Deutschland: Prüfung am Patienten. In: Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA), des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ) und der Chirurgischen Arbeitsgemeinschaft Lehre (CAL). Frankfurt am Main, 25.-28.09.2019. Düsseldorf: German Medical Science GMS Publishing House; 2019. DocP-05-03. DOI: 10.3205/19gma287
26. Mihaljevic AL, Schmidt J, Mitzkat A, Probst P, Kenngott T, Mink J, Fink CA, Ballhausen A, Chen J, Cetin A, Murrmann L, Müller G, Mahler C, Götsch B, Trierweiler-Hauke B. Heidelberger interprofessionelle Ausbildungsstation (HIPSTA): a practice- and theory-guided approach to development and implementation of Germany's first interprofessional training ward. *GMS J Med Educ*. 2018;35(3):Doc33. DOI: 10.3205/zma001179
27. Post R, Jonietz A, Selgert L, Lux R, Mihaljevic AL, Jünger J. Entwicklung, Testung und Validierung eines Bewertungsbogens zur Beurteilung laienverständlicher Patientenbriefe. In: Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA), des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ) und der Chirurgischen Arbeitsgemeinschaft Lehre (CAL). Frankfurt am Main, 25.-28.09.2019. Düsseldorf: German Medical Science GMS Publishing House; 2019. DocP-05-02. DOI: 10.3205/19gma286
28. Weins C. Uni- und bivariate deskriptive Statistik. In: Wolf C, Best H, editors. *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS-Verlag für Sozialwissenschaften; 2010. p.65-89. DOI: 10.1007/978-3-531-92038-2_4
29. Field S, Egan R, Beesley T. Applying G-Theory and Multivariate G-Analysis to improve Clinical Data quality and performance assessment accuracy. OHSE Working Paper. San Francisco, CA: Academia.edu; 2017. Zugänglich unter/available from: https://www.academia.edu/37118667/_Applying_G-Theory_and_Multivariate_G-Analysis_to_improve_Clinical_Data_quality_and_performance_assessment_accuracy_2017_OHSE_Working_Paper
30. Schmolck P. Begleittext: Methoden der Reliabilitätsschätzung. München: Universität der Bundeswehr München; 2007. Zugänglich unter/available from: <https://dokumente.unibw.de/pub/bscw.cgi/1787978>
31. Cardinet J, Tourneur Y, Allal L. Extension of generalizability theory and its applications in educational measurement. *J Educ Measurement*. 1981;18(4):183-204. DOI: 10.1111/j.1745-3984.1981.tb00852.x
32. Bühl A. SPSS 16. Einführung in die moderne Datenanalyse. 11th ed. München: Pearson-Studium; 2008.
33. Döring N, Bortz J. Datenerhebung. In: Döring N, Bortz J, editors. *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer; 2016. p.321-577. DOI: 10.1007/978-3-642-41089-5_10
34. Bortz J, Lienert GA, Boehnke K. Verteilungsfreie Methoden in der Biostatistik. 3rd ed. Heidelberg: Springer Medizin Verlag; 2008.
35. Rammstedt B. Reliabilität, Validität, Objektivität. In: Wolf C, Best H, editors. *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS-Verlag für Sozialwissenschaften; 2010. p.239-258. DOI: 10.1007/978-3-531-92038-2_11
36. Bittner A, Bittner J, Jonietz A, Dybowski C, Harendza S. Translating medical documents improves students' communication skills in simulated physician-patient encounters. *BMC Med Educ*. 2016;16:72. DOI: 10.1186/s12909-016-0594-4
37. Operation Karriere. Unsere Vision ist es, Arzt und Patient auf Augenhöhe zu bringen. Köln: Deutscher Ärzteverlag GmbH; 2017. Zugänglich unter/available from: <https://www.operation-karriere.de/karriereweg/bewerbung-berufstart/unsere-vision-ist-es-arzt-und-patient-auf-augenhoehe-zu-bringen.html>
38. Kühnel SM, Krebs D. Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. 4th ed. Reinbek: Rowohlt; 2007.
39. Schmidt RC. Managing Delphi surveys using nonparametric statistical techniques. *Decision Sci*. 1997;28(3):763-774. DOI: 10.1111/j.1540-5915.1997.tb01330.x

Corresponding author:

Dr. med. Lena Selgert

Institut für medizinische und pharmazeutische
Prüfungsfragen (IMPP), Postfach 2518, D-55015 Mainz,
Germany
LSelgert@impp.de

Please cite as

Selgert L, Bender B, Hinding B, Federmann A, Mihaljevic AL, Post R, Jonietz A, Norcini J, Tekian A, Jünger J. Development, testing and generalizability of a standardized evaluation form for the assessment of patient-directed reports in the new final medical licensing examination in Germany. *GMS J Med Educ*. 2021;38(3):Doc71. DOI: 10.3205/zma001467, URN: urn:nbn:de:0183-zma0014672

This article is freely available from

<https://www.egms.de/en/journals/zma/2021-38/zma001467.shtml>

Received: 2020-03-31

Revised: 2020-08-10

Accepted: 2020-09-21

Published: 2021-03-15

Copyright

©2021 Selgert 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, Testung und Generalisierbarkeit eines standardisierten Bewertungsbogens zur Beurteilung von patientenverständlichen Berichten im künftigen abschließenden Staatsexamen des Medizinstudiums in Deutschland

Zusammenfassung

Hintergrund: Ärzt*innen erklären ihren Patient*innen Diagnosen und Therapien oft nicht ausreichend verständlich und nachvollziehbar. Eine Verbesserung der Arzt-Patienten-Kommunikation ist daher dringend geboten. Die derzeitige medizinische Ausbildung inkl. der Prüfungen konzentriert sich auf die Vermittlung von Kompetenzen in der mündlichen Kommunikation, während die schriftliche Kommunikation mit Patient*innen bislang nur wenig Beachtung findet.

Zur optimalen Vorbereitung der Medizinstudierenden auf den Berufsstart wurde das abschließende Staatsexamen des Medizinstudiums in Deutschland vom Institut für medizinische und pharmazeutische Prüfungsfragen im Sinne von „Assessment drives learning“ neugestaltet. Künftig sollen die Prüflinge als Teil des Entlassmanagements einen Bericht für die Patient*innen erstellen, der diesen in patientenverständlicher Sprache alle relevanten Informationen über ihren Krankenhausaufenthalt zugänglich macht.

Zielsetzung: Hierfür wurde ein standardisierter Bewertungsbogen entwickelt. Mit diesem kann ein formatives und summatives Feedback zu den Kompetenzen in schriftlicher Kommunikation gegeben werden. Die Anwendbarkeit dieses Bogens sowie die Testgütekriterien, insbesondere die Reliabilität, wurden getestet.

Methodik: In einem Expertenkonsensusverfahren wurde ein erster Entwurf dieses Bewertungsbogens entwickelt. Dieser Bogen wurde nach einem ersten Probelauf mit patientenverständlichen Berichten, die von Studierenden im letzten Jahr ihres Medizinstudiums verfasst wurden, überarbeitet. Anschließend wurden 21 patientenverständliche Berichte von 14 verschiedenen Prüfenden bewertet. Die Reliabilität wurde durch die Berechnung des Generalisierungskoeffizienten und durch die Analyse der Inter-Rater-Reliabilität getestet.

Ergebnisse: Im Rahmen der ersten Testung der Bewertung von patientenverständlichen Berichten konnte die Praktikabilität und der Nutzen des Bewertungsbogens als Instrument zur Beurteilung der spezifischen schriftlichen Kommunikationsfähigkeiten von Studierenden festgestellt werden. Die Analysen der Inter-Rater-Reliabilität zeigten, dass sich der Grad der Übereinstimmung in den Bewertungen je nach Vorkenntnissen der Prüfenden im Bereich der patientenverständlichen Kommunikation teilweise unterschied. Der berechnete G-Koeffizient weist auf eine hohe Reliabilität des Bewertungsbogens hin. Die inhaltliche Validität des Bewertungsbogens konnte durch die umfassende medizinische Expertise im Entwicklungsprozess sichergestellt werden.

Schlussfolgerung: Die Bewertung der Kommunikation mit Patient*innen als Teil der neu entwickelten Prüfung am Patienten oder an der Patientin bedeutet einen Zugewinn für die medizinische Ausbildung. Regelmäßige formative Prüfungen und Feedback anhand des Bewertungsbogens können im Laufe des Studiums eingesetzt werden, um die schriftlichen Kommunikationskompetenzen künftiger Ärzt*innen zu verbessern.

Lena Selgert¹
Bernd Bender¹
Barbara Hinding¹
Aline Federmann¹
André L. Mihaljevic²
Rebekka Post³
Ansgar Jonietz³
John Norcini⁴
Ara Tekian⁵
Jana Jünger¹

1 Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP), Mainz, Deutschland

2 Universitätsklinikum Heidelberg, Klinik für Allgemein-, Viszeral- und Transplantationschirurgie, Heidelberg, Deutschland

3 "Was hab' ich?" gGmbH, Dresden, Deutschland

4 SUNY Upstate Medical University, Department of Psychiatry, New York, USA

5 University of Illinois at Chicago, College of Medicine, Illinois, USA

Darüber hinaus ist zu erwarten, dass das Verstehen von Diagnosen und Therapien sowie eine vertrauensvolle Arzt-Patienten-Beziehung Patient*innen zu mehr Teilhabe am medizinischen Entscheidungsprozess befähigt. Dies könnte künftig auch Entlassungsfehlern vorbeugen. Für den einheitlichen Gebrauch des Bewertungsbogens ist die Implementierung standardisierter Prüferschulungen erforderlich.

Schlüsselwörter: Kommunikation, Bildung, Patientenbeteiligung

1. Einleitung

Nach dem "Patientenrechtegesetz" des deutschen Bürgerlichen Gesetzbuches hat jeder*jede Patient*in das Recht, umfassend informiert zu werden [1]. Dennoch zeigen mehrere Studien, dass Ärzt*innen ihren Patient*innen Diagnosen und Therapien oft nicht verständlich und angemessen erklären. Daher ist eine Verbesserung der Arzt-Patienten-Kommunikation unerlässlich: In 22% der Fälle erhalten die Patient*innen unverständliche Antworten auf ihre Fragen und in 29% unverständliche Erklärungen ihrer Untersuchungsergebnisse durch ihre Ärzt*innen. Infolgedessen fühlen sich 39% der Patient*innen mit ihren Sorgen und Ängsten allein gelassen [2].

Wie eine zwischen Oktober 2012 und September 2013 durchgeführte Studie ergab, wurden bis zu 33% der Entlassungsfehler durch mangelhafte Kommunikation des medizinischen Fachpersonals verursacht [3]. Hierbei handelte es sich um Medikationsfehler, schlechte Wundversorgung oder unzureichende Ernährung, die zu Rehospitalisierung, lebensbedrohlichen Situationen, vermeidbaren oder unnötigen medizinischen Leistungen und Prozeduren sowie zu zusätzlichem Aufwand für die Pflegedienste und erhöhten Kosten für das Gesundheitssystem führten. Die Bereitstellung sowohl umfassender als auch schriftlicher Informationen für Patient*innen ist wichtig, um sowohl deren Therapietreue zu steigern als auch präventive Maßnahmen durchzuführen [2], [4], [5], [6], [7]. Insbesondere das Vermeiden medizinischen Fachjargons wird als sinnvolle Strategie hervorgehoben, um Patient*innen in ihrem Entscheidungsprozess zu stärken [8].

Um die Arzt-Patienten-Kommunikation zukünftiger Ärzt*innen zu verbessern, entwickelten Jünger et al. ein longitudinales Mustercurriculum Kommunikation zur Implementierung in das Medizinstudium [9], [10]. Auf dessen Grundlage können die wesentlichen Lernziele der Arzt-Patienten-Kommunikation in die medizinische Lehre und Prüfung integriert werden [11].

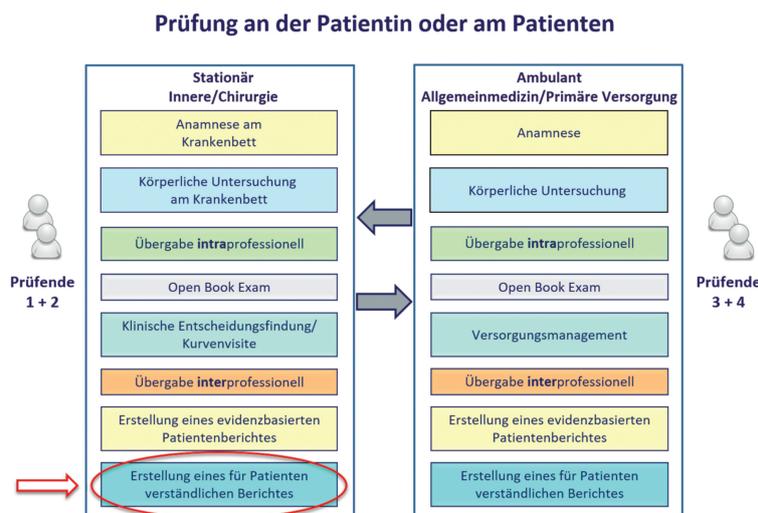
Im Sinne von „Assessment drives learning“ [12] wurde das künftige abschließende Staatsexamen des Medizinstudiums in Deutschland vom Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP) im Hinblick auf eine größere Patientenorientierung neugestaltet. Zur Beurteilung der kommunikativen Kompetenzen werden an medizinischen Fakultäten häufig „objective structured clinical examinations“ (OSCE) eingesetzt, die mithilfe von Simulationspersonen schwerpunktmäßig verbale Kommunikationsfähigkeiten erfassen [13]. Um während der neuen Staatsexamensprüfung eine authentische Situation zu schaffen, findet die hierbei vorgesehene arbeitsplatzbasierte Prüfung an realen Patient*innen auf einer chirurgischen oder internistischen Station sowie in einer allgemeinärztlichen Praxis statt (siehe Abbildung 1). Als Teil des verbesserten Entlassmanagements erstellen die Prüflinge einen evidenzbasierten Patientenbericht für den*die weiterbehandelnden Arzt*Ärztin sowie einen Bericht für die Patient*innen selbst, der leicht verständ-

lich ist und ihnen alle relevanten Informationen liefert [14].

Die Verwendung einer einfachen Sprache ist eine der häufigsten Strategien, die von Ärzt*innen, Pflegekräften und Apotheker*innen zur Verbesserung der Gesprächsführung mit ihren Patient*innen eingesetzt wird [15]. Earl et al. zeigten die Auswirkungen eines Gesundheitskompetenzmoduls auf die Verbesserung des schriftlichen Patientenaufklärungsmaterials der Studierenden in den Bereichen Lesbarkeit, Inhalt, Rechenleistung, Statistik und Konzepte der Patientenaktivität. Die Vereinfachung der medizinischen Sprache gestaltete sich allerdings schwierig [16].

Zusammenfassend lässt sich sagen, dass es bereits eine Reihe internationaler Studien zur Arzt-Patienten-Kommunikation im Allgemeinen und der schriftlichen Kommunikation im Besonderen gibt [17], [18], [19], [20], [21]. Bisher befassten sich diese Studien vor allem mit der Beziehung zwischen Gesundheitskompetenz der Patient*innen und Kommunikation, der Lesbarkeit und Verständlichkeit schriftlicher Patienteninformationsmaterialien sowie mit dem Nutzen schriftlicher Kommunikationsstrategien. Eine größere randomisierte, kontrollierte Studie zum Nutzen laienverständlicher Patientenberichte wurde bereits von der Initiative „Was hab' ich?“ gGmbH mit 417 Patient*innen durchgeführt. Hierbei stellten die Ärzt*innen von „Was hab' ich?“ den Patient*innen nach der Entlassung aus dem Krankenhaus einen leicht verständlichen Patientenbericht zur Verfügung. Diese Studie zeigte signifikante Auswirkungen der Patientenberichte auf das Verständnis der Patient*innen von Untersuchungsergebnissen, Medikamentenindikationen und -verschreibungen [22]. Die in den Berichten verwendete Sprache zeichnete sich durch einfache Wörter, kurze, vollständige und einfache Sätze sowie durch eine positive Sprache und das Vermeiden medizinischer Fachausdrücke aus. Zudem wurden relevante Hintergrundinformationen bereitgestellt und auf einen inhaltlich logischen Aufbau des Texts geachtet [23], [24]. Auf der Grundlage dieser Kritisierensammlung für die Erstellung eines Berichts in patientenverständlicher Sprache entwarf „Was hab' ich?“ eine Vorlage für die Erstellung dieser Berichte durch Medizinstudierende. Bisher gibt es jedoch kein Bewertungsinstrument zum Verfassen für patientenverständliche Berichte. Ziel dieser Studie war daher die Entwicklung und Testung eines standardisierten Bewertungsbogens für die formative und summative Beurteilung von Berichten, der die wichtigen Aspekte des patientenverständlichen Schreibens umfasst. Basierend auf realen anstelle von simulierten Patient*innen und Situationen muss dieser Bogen individuell verwendet werden und alle Konsultationsanlässe und Erkrankungen abdecken können.

Zur Entwicklung dieses Bewertungsbogens wurde die Kritisierensammlung der „Was hab' ich?“ gGmbH herangezogen, eine Literaturanalyse durchgeführt und Expertenmeinungen eingeholt. Geprüft wurden die Anwendbarkeit des Bogens und die Testgütekriterien, insbesondere die Reliabilität.



Unter der Aufsicht von zwei Prüfenden werden in acht Schritten die ärztlichen Basiskompetenzen erfasst, beginnend mit einer vollständigen Anamnese und körperlichen Untersuchung zuvor unbekannter Patient*innen. Die erhobenen Informationen werden strukturiert an die Prüfenden übergeben. Anschließend beantworten die Studierenden eine auf die Patient*innen bezogene, konkrete Fragestellung nach dem PICO-Schema (Patient Population – Intervention – Comparison – Outcome) und Durchführung einer wissenschaftlichen Literaturrecherche (Open Book Exam) und füllen eine standardisierte Patientenkurve aus. Die Patientenkurve wird den Prüfenden präsentiert und dient der Reflexion klinischer Entscheidungen. Anschließend erfolgt zur Demonstration der interprofessionellen Kompetenzen eine Patientenübergabe an die zuständige Person aus dem interprofessionellen Behandlungsteam. Zum Abschluss erstellen die Studierenden einen evidenzbasierten Patientenbericht sowie einen für Patient*innen verständlichen Bericht, der ihnen selbst alle wichtigen Informationen zum Krankenhausaufenthalt vermittelt [14].

Abbildung 1: Arbeitsplatzbasierte Prüfung an realen Patient*innen auf einer chirurgischen oder internistischen Station oder in der ambulanten Versorgung.

2. Methodik

2.1. Entwicklung eines standardisierten Bewertungsbogens für patientenverständliche Berichte

Zur Beurteilung der Qualität schriftlicher Kommunikation Medizinstudierender in patientenverständlichen Berichten wurde auf der Grundlage einer Literaturanalyse und einer Kriteriensammlung für das patientenverständliche Schreiben ein erster Entwurf eines Bewertungsbogens durch die „Was hab' ich?“ gGmbH und das IMPP erstellt [23], [24]. Im August 2018 erarbeitete eine Gruppe von 27 medizinischen Expert*innen aus acht deutschen Fakultäten, bestehend aus Fachärzt*innen für Allgemeinmedizin, Innere Medizin, Anästhesie, Psychiatrie, Chirurgie, Psychosomatik und Psychotherapie, Psycholog*innen, die alle am Weiterbildungsstudiengang „Master of Medical Education“ (MME) teilgenommen haben, sowie fünf Studierenden hieraus gemeinsam einen Bewertungsbogen: Der erste Entwurf wurde zunächst in einer Kleingruppe von sieben Expert*innen überarbeitet und anschließend in der Gesamtgruppe diskutiert und konsensual beschlossen.

Im Oktober 2018 wurde dieser Bewertungsbogen mit insgesamt zehn Prüflingen an der Klinik für Allgemein-, Viszeral- und Transplantationschirurgie des Universitätsklinikums Heidelberg getestet [25]. Die Prüflinge befanden sich im vierten bis sechsten Jahr ihres Medizinstudiums. Jeder Prüfling schrieb einen patientenverständlichen

Bericht basierend auf einer standardisierten Vorlage. Diese Berichte wurden anschließend von elf Ärzt*innen und drei Studierenden begutachtet, die an der Entwicklung des Bewertungsbogens im August 2018 mitgewirkt hatten. Basierend auf diesen Erfahrungen wurde der Bewertungsbogen von den teilnehmenden Ärztinnen der Initiative „Was hab' ich?“ gGmbH und des IMPP weiterentwickelt.

2.2. Testung und Überarbeitung des Bewertungsbogens

Im Januar und Februar 2019 wurde der überarbeitete Bewertungsbogen anhand von 21 patientenverständlichen Berichten, die von Studierenden im letzten (sechsten) Jahr ihres Medizinstudiums (PJ) auf der Heidelberger interprofessionellen Ausbildungsstation (HIPSTA) des Universitätsklinikums Heidelberg verfasst wurden, umfassend getestet. Auf der HIPSTA behandeln Medizinstudierende und Pflegeauszubildende vier Wochen lang gemeinsam Patient*innen unter der Aufsicht von medizinischen und pflegerischen Lernbegleiter*innen [26]. Ein Teil der praktischen Ausbildung besteht darin, zusätzlich zum konventionellen Entlassbericht an die weiterbehandelnden Ärzt*innen, auch patientenverständliche Berichte zu verfassen. Die PJ-Studierenden erhielten von der Initiative „Was hab' ich?“ gGmbH hierzu im Vorfeld ein strukturiertes Training im patientenverständlichen Schreiben. Hierbei wurden das sichere Erkennen und Vermeiden medizi-

nischer Begriffe, die Erläuterung von Hintergrundinformationen und ein einfach strukturierter Schreibstil geübt. Diese 21 patientenverständlichen Berichte wurden von zwei Prüfergruppen, bestehend aus jeweils zwei Ärztinnen der Initiative „Was hab' ich?“ gGmbH, bewertet. Die erste Gruppe verwendete den entwickelten Bewertungsbogen mit den detaillierten Unterpunkten. Die zweite Gruppe, die die Berichte auswertete, konzentrierte sich hingegen nur auf die drei Hauptkriterien ohne Kenntnis der Unterpunkte [27].

2.3. Evaluierungsstudie

Nach der Überarbeitung des Bewertungsbogens auf Basis der ersten Testung auf der HIPSTA wurden die 21 patientenverständlichen Berichte erneut von insgesamt 14 Prüfenden bewertet. Diese Prüfenden setzten sich nun zusammen aus Ärztinnen der Initiative „Was hab' ich?“ gGmbH mit Erfahrung im patientenverständlichen Schreiben und aus Ärzt*innen des IMPP sowie Allgemeinmediziner*innen, da letztere die Ärzt*innen sind, die in der Regel Entlassberichte erhalten. Der entsprechende Entlassbericht für die weiterbehandelnden Ärzt*innen wurde zum Vergleich zur Verfügung gestellt. Die für die statistische Analyse verwendete Stichprobe umfasste die vorhandenen Bewertungen aller Prüfenden. Da nicht alle Prüfenden alle Berichte bewerteten, ergab sich eine Stichprobe von N=205 Bewertungen. Die Stichprobe für die Analysen zur Inter-Rater-Reliabilität bestand aus den Bewertungen der neun Prüfenden, die elf der 21 Berichte vollständig bewertet hatten. Da keine persönlichen Daten ausgewertet wurden, war keine Zustimmung der Ethikkommission erforderlich.

2.4. Testqualitätskriterien und statistische Methoden

Statistisch wurden zunächst deskriptive Kennzahlen, die Mittelwerte und Standardabweichungen, für jede der drei Kategorien des Bewertungsbogens einzeln analysiert. Darüber hinaus wurden bivariate Korrelationen zwischen den Kategorien berechnet. Hierfür wurde der Korrelationskoeffizient nach Pearson angewendet [28].

Die Reliabilität ist gegeben, wenn das jeweilige Instrument keine Messfehler erzeugt. Sie kann angenommen werden, wenn ein hoher Grad an Interkorrelation zwischen den einzelnen Teilen eines Messinstruments besteht. Die Reliabilität wurde durch die Berechnung des Generalisierbarkeits- (G-) Koeffizienten analysiert [29]. Es wurde gegen Cronbachs Alpha entschieden, um eine Überschätzung der Reliabilität aufgrund von „Übungseffekten“ bei Prüfenden nach mehrmaligem Einsatz des Bewertungsbogens zu vermeiden [30]. Daher wurden Varianzkomponenten für die Faktoren Prüfende, Bericht und Hintergrund der Prüfenden („Was hab' ich?“ vs. IMPP und Hausärzt*innen) bezüglich der erreichten Punktzahl bei allen Bewertungen basierend auf der Generalisierbarkeitstheorie berechnet [29], [31]. Dies half, mögliche Quellen

für Messfehler in den Bewertungen der Berichte zu identifizieren [32]. Die relative Fehlervarianz wurde auf der Grundlage der berechneten Varianzkomponenten bestimmt. Mit dieser relativen Fehlervarianz konnte der G-Koeffizient berechnet werden. Der G-Koeffizient schätzt, ob die Ergebnisse auf die Population übertragbar sind oder ob der Interaktionseffekt zwischen den Faktoren und den Prüflingen nur auf die Studienstichprobe bezogen werden kann. Ein G-Koeffizient von eins zeigt an, dass die verfügbaren Daten und Ergebnisse perfekt auf alle Bewertungen außerhalb der Studie übertragen werden können. Ein hoher Wert für diesen Koeffizienten deutet also auf eine hohe Reliabilität hin [29].

Die Bewertungsübereinstimmung in Form der Inter-Rater-Reliabilität [33] wurde auf der Basis einer Stichprobe von neun Prüfenden berechnet, die elf der 21 Berichte vollständig bewerteten. Die anderen Prüfenden bewerteten einen Teil der Berichte. Die Inter-Rater-Reliabilität wurde für jede der drei Kategorien in den elf Berichten getrennt berechnet. Dies basierte auf dem Koeffizienten Kendall W [34]. Die Bewertungsskala des entwickelten Bewertungsbogens stieg von null bis fünf Punkten auf. Da mindestens ein ordinales Skalenniveau angenommen werden konnte, erschien Kendall W als der geeignete Koeffizient, im Gegensatz z. B. zum Fleiss Kappa-Koeffizienten, welcher kategoriale Daten voraussetzt [34]. Ziel war es, festzustellen, ob die Prüfenden innerhalb einer Bewertungskategorie in ihren Bewertungen konsistent waren oder nicht.

Darüber hinaus wurde die inhaltliche Validität des Bewertungsbogens durch die umfassende medizinische Expertise im Entwicklungsprozess sichergestellt. Dies bedeutet, dass die entwickelten Kategorien des Bewertungsbogens den beabsichtigten Inhalt repräsentieren [35].

Um Beeinflussungen in der Bewertung der patientenverständlichen Berichte zu vermeiden, kannten sich die an der Testung beteiligten Prüfenden und Studierenden nicht persönlich.

3. Ergebnisse

3.1. Entwicklung und erste Testung des Bewertungsbogens

Der erste Entwurf des Bewertungsbogens enthielt die Kategorien „Inhaltsauswahl“, „Medizinische Korrektheit“, „Struktur und Syntax des Berichts“, „Sprachliche Gestaltung“ und „Grammatik“. In jeder Kategorie konnten null bis fünf Punkte vergeben werden. Die Kategorien wurden in der Gesamtbewertung mit zehn bis 30% gewichtet.

Den Ergebnissen der ersten Testung folgend wurden die Hauptbewertungskriterien zu drei Kategorien zusammengefasst: „Inhaltsauswahl und medizinische Korrektheit“, „Übersetzung medizinischer Fachsprache in laienverständliche Sprache“ und „Leicht verständliche Sprache“. Jede der drei Hauptbewertungskategorien wurde mit spezifizierten Unterpunkten detaillierter dargestellt. In jeder der drei Kategorien konnten null bis fünf Punkte vergeben

werden. Die Kategorie „Inhaltsauswahl und medizinische Korrektheit“ ging mit 40% in die Gesamtbewertung ein. Die anderen beiden Kategorien erhielten jeweils eine Gewichtung von 30%.

Diese Version des Bewertungsbogens wurde gemäß den Testungen auf der HIPSTA überarbeitet und spezifiziert. Die zweite Kategorie wurde in „Laienverständliche Sprache und Hintergrundinformation“ umbenannt. Die Anzahl der Unterpunkte wurde in der Kategorie „Inhaltsauswahl und medizinische Korrektheit“ von ursprünglich acht zu fünf Unterpunkten zusammengefasst.

3.2. Getestete Version des Bewertungsbogens

Ein standardisierter Bewertungsbogen für patientenverständliche Berichte konnte erfolgreich mit drei Hauptbewertungskategorien entwickelt werden. „Inhaltsauswahl und medizinische Korrektheit“ wurde mit 40 Prozent und „Laienverständliche Sprache und Hintergrundinformation“ sowie „Leicht verständliche Sprache“ mit jeweils 30 Prozent in der Gesamtbewertung berücksichtigt. Die Unterpunkte dienen zur Spezifizierung der Bewertungskategorien (siehe Abbildung 2).

3.3. Evaluation

3.3.1. Deskriptive Kennzahlen und Korrelationen

Aus der Stichprobe von 205 individuellen Bewertungen der Prüfenden wurden Mittelwerte, Standardabweichung und paarweise Korrelationen zwischen den drei Kategorien berechnet.

Folgend wird ein Überblick über die Mittelwerte sowie die Standardabweichung für jede der drei Kategorien gegeben (siehe Tabelle 1).

Alle drei Kategorien wiesen nahezu identische Mittelwerte und Standardabweichungen auf. Somit hat jeder der 14 Prüfenden im Durchschnitt die gleiche Punktzahl in jedem der 21 Berichten vergeben. Hieraus folgt, dass keine der Kategorien deutlich besser oder schlechter bewertet als die andere.

Da sich aus diesen deskriptiven Zahlen bereits mögliche Zusammenhänge abzeichneten, wurden die bivariaten Korrelationen zwischen den Kategorien berechnet (siehe Tabelle 2).

Die Werte der paarweisen Korrelation zeigten einen mittleren bis hohen positiven Zusammenhang zwischen den drei Kategorien, wobei alle Korrelationen hochsignifikant waren (p -Wert $< 0,001$). Die stärkste Korrelation wurde zwischen den Kategorien „Laienverständliche Sprache und Hintergrundinformation“ und „Leicht verständliche Sprache“ gefunden ($r=0,61$). Im Vergleich hierzu korrelierten „Inhaltsauswahl und medizinische Korrektheit“ und „Laienverständliche Sprache und Hintergrundinformation“ ($r=0,45$) sowie „Inhaltsauswahl und medizinische Korrektheit“ und „Leicht verständliche Sprache“ ($r=0,31$) auf einem mittleren Niveau.

3.3.2. Reliabilität

Die Interrater-Reliabilität der elf Berichte wurde für jede der drei Bewertungskategorien separat ausgewertet. Sie wurde getrennt nach Prüfenden der „Was hab' ich?“ gGmbH und Prüfenden des IMPP und der Allgemeinmedizin berechnet, um die Bewertungen beider Gruppen vergleichen zu können (siehe Tabelle 3).

Die Übereinstimmungsrate in der ersten Kategorie war bei den Prüfenden von „Was hab' ich?“ im Vergleich zu den Prüfenden des IMPP und den Allgemeinmediziner*innen deutlich höher. Im Vergleich hierzu war die Übereinstimmung der Bewertungen für die anderen beiden Kategorien in beiden Gruppen auf einem mittleren Niveau. Der berechnete G-Koeffizient lag bei 0,72 für alle 205 Bewertungen. Dieser eher hohe Wert zeigt, dass die Bewertungen der patientenverständlichen Berichte nicht nur auf die Stichprobe begrenzt sind, sondern auch auf Bewertungen außerhalb der Studie übertragen werden können.

3.3.3. Überarbeitung des Bewertungsbogens auf Grundlage der Evaluation

Der Bewertungsbogen wurde anhand der Evaluationsergebnisse leicht überarbeitet: Die Kategorie „Laienverständliche Sprache und Hintergrundinformationen“ wurde in „Vermittlung von Hintergrundinformationen und patientenverständlicher Umgang mit Fachbegriffen“ spezifiziert. Zur besseren Unterscheidung der beiden Kategorien wurde die Kategorie „Leicht verständliche Sprache“ in „Patientenverständlicher Sprachstil, Lesbarkeit und Alltagssprache“ umbenannt.

Die Erklärung des Unterpunkts „Ihre Medikamente“ wurden um „Erklärt das Einnahmeschema, gibt Einnahmehinweise“ und „Weist ggf. auf relevante Interaktionen und/ oder unerwünschte Wirkungen hin“ ergänzt. Zudem wurde die Erklärung „Gibt Verhaltensempfehlungen“ beim Unterpunkt „Die nächsten Schritte“ um die Punkte „Hygiene, Wundpflege, Ernährung, Bewegung, Trinkmenge, Nikotin“ erweitert.

4. Diskussion

Mittels eines mehrstufigen Konzeptions- und Überarbeitungsprozess wurde ein standardisierter Bewertungsbogen zur Beurteilung von patientenverständlichen Berichten erfolgreich entwickelt. Dies ist eine bemerkenswerte Verbesserung der Ausbildung der Studierenden im Bereich der Arzt-Patienten-Kommunikation.

Eine Studie der „Was hab' ich?“ gGmbH zeigte, dass Medizinstudierende, die ein Training zur schriftlichen Kommunikation absolviert haben und regelmäßig Befunde übersetzen, im Gespräch mit standardisierten Personen in einem simulierten Arzt-Patienten-Kontakt bessere Erklärungen verwenden als ungeschulte Studierende [36]. Dies entspricht auch dem Selbstverständnis der Studierenden, die bei der „Was hab' ich?“ gGmbH arbei-



		Punkte
Inhaltliche Auswahl und fachliche Korrektheit (40%)		0-5
- wählt die für den Patienten oder die Patientin relevanten Informationen aus - gibt die Informationen fachlich korrekt wieder		
Der Grund für Ihren Krankenhausaufenthalt	- Erklärt den Einweisungsgrund in wenigen Sätzen inkl. Beschwerden, bisheriger Diagnostik und Therapie	
Ihre Krankengeschichte und Ihre weiteren Erkrankungen	- Beschreibt wichtige Ergebnisse der Anamnese-Erhebung und der körperlichen Untersuchung - Erklärt wichtige Nebendiagnosen	
Der Verlauf Ihres Krankenhaus-Aufenthalts	- Erklärt schlüssig den Verlauf mit durchgeführter Diagnostik und Therapie - Erläutert ggf. Komplikationen	
Ihre Medikamente	- Erklärt jeweils die Indikation/gewünschte Wirkung - Erklärt das Einnahmeschema	
Die nächsten Schritte	- Gibt Informationen zu weiterer geplanter Diagnostik und Therapie inkl. Termine, Rehabilitation etc. - Gibt Verhaltensempfehlungen (Hygiene, Wundpflege, Ernährung) und erklärt konkretes Vorgehen bei Komplikationen	
Laiensprache und Hintergrundinformation (30%)		0-5
- Findet und nutzt im gesamten Text gute Entsprechungen für Fachwörter - Erklärt schwierige Wörter und gibt angemessene Hintergrundinformationen - Erklärt relevante anatomische und physiologische Zusammenhänge - Nutzt nur allgemein geläufige Abkürzungen		
Leicht verständliche Sprache (30%)		0-5
- Spricht den Patienten oder die Patientin direkt an - Formuliert ganze, kurze und einfach aufgebaute Sätze - Verwendet einfache Wörter aus dem Alltag - Vermeidet gehobenen Sprachstil (wie z.B. Substantivierungen, Konjunktiv, Passiv, Imperfekt) - Schreibt orthographisch und grammatikalisch korrekt		
Gesamtbewertung		/100%

Abbildung 2: Getesteter Bewertungsbogen für die Beurteilung von patientengerichteten Berichten

Tabelle 1: Deskriptive Statistiken der drei Bewertungskategorien

Kategorie	MW (SD)
Inhaltliche Auswahl und technische Korrektheit	3,8 (0,99)
Laiensprache und Hintergrundinformation	3,8 (0,97)
Leicht verständliche Sprache	3,7 (0,95)

Deskriptive Statistiken (Mittelwert (MW) und Standardabweichung (SD)) der drei Kategorien auf Basis der Stichprobe von insgesamt N=205 Bewertungen.

Tabelle 2: Bivariate Korrelationen zwischen den drei Bewertungskategorien

Kategorien	Pearson's r
Inhaltliche Auswahl und technische Korrektheit - Laiensprache und Hintergrundinformation	0,5***
Inhaltliche Auswahl und technische Korrektheit - Leicht verständliche Sprache	0,31***
Laiensprache und Hintergrundinformation - Leicht verständliche Sprache	0,61***

Bivariate Korrelationen anhand des Pearson's Korrelationskoeffizienten zwischen den drei Bewertungskategorien auf Basis der Stichprobe von insgesamt n = 205 Bewertungen. Die Einteilung der Werte des Pearson-Korrelationskoeffizienten nach Kühnel und Krebs [36] sind: 0,00<r<0,05 = kein Zusammenhang; 0,05<r<0,20 = geringer Zusammenhang; 0,20<r<0,50 = mittlerer Zusammenhang; 0,50<r<0,70 = hoher Zusammenhang; r>0,70 = sehr hoher Zusammenhang. Das verwendete Signifikanzniveau betrug p ≤0,05. Legende: n. s. = nicht signifikant; * = p-Wert ≤0,05; ** = p-Wert <0,01; *** = p-Wert <0,001.

ten: Sie sind sich einig, dass die schriftliche Übersetzung von fachsprachlichen Entlassberichten in laienverständliche Patientenberichte zu einer Steigerung ihrer Kompetenzen in patientenverständlicher Kommunikation führt [37].

In der hier vorliegenden Studie erstellten Studierende, die sich im letzten Jahr ihrer medizinischen Ausbildung befanden, patientenverständliche Berichte. Die Testung

des neu entwickelten Bewertungsbogens, der bei 21 dieser patientenverständlichen Berichte angewendet wurde, konnte die Praktikabilität sowie den Nutzen dieses Instruments zur Beurteilung der spezifischen schriftlichen Kommunikationsfähigkeiten von Studierenden zeigen. Durch die Einbindung verschiedener medizinischer Expert*innen in den Entwicklungsprozess wurde sichergestellt, dass die einzelnen Bewertungskategorien die wichtigsten Schritte beim Verfassen eines patientenver-

Kategorie	Prüfende von „Was hab' ich?“ (N=5)	Prüfende des IMPP und aus der Allgemeinmedizin (N=4)
	Kendall W	Kendall W
Inhaltliche Auswahl und technische Korrektheit	0,82	0,35
Laiensprache und Hintergrundinformation	0,51	0,58
Leicht verständliche Sprache	0,64	0,65

Inter-Rater-Reliabilität der elf patientenverständlichen Berichte. Kendall Konkordanzkoeffizient W: 0 = keine Übereinstimmung, 1 = perfekte Übereinstimmung. Die Einteilung der Werte für Kendall W nach R. C. Schmidt [37] sind: 0,1 – sehr schwache Übereinstimmung; 0,3 – schwache Übereinstimmung; 0,5 – moderate Übereinstimmung; 0,7 – starke Übereinstimmung; 0,9 – extrem starke Übereinstimmung.

Tabelle 3: Inter-Rater-Reliabilität für jede Bewertungskategorie

ständlichen Berichtes abbilden. Die Implementierung von spezifischen Unterpunkten unterstützt die Prüfenden bei der Interpretation der Kategorien.

Zwischen den drei Kategorien konnten mittlere und hohe Korrelationen beobachtet werden. Vor allem die hohe Korrelation zwischen „Laienverständliche Sprache und Hintergrundinformation“ und „Leicht verständliche Sprache“ ist ein Indikator für die gute Reliabilität des Bewertungsbogens. Dagegen sind die mittelstarken Korrelationen zwischen „Inhaltsauswahl“ und „Laienverständliche Sprache und Hintergrundinformation“ sowie „Leicht verständliche Sprache“ ein Indiz für die Unabhängigkeit der Bewertungen von Inhalt und Sprache. Jene Studierenden, die nach Meinung der Prüfenden in den Patientenberichten eine angemessene Laiensprache verwendeten und die Hintergrundinformationen gut kommunizierten, waren auch in der Lage, in einer leicht verständlichen Sprache zu schreiben. Im Gegensatz dazu standen eine angemessene Auswahl des Inhalts und die fachliche Korrektheit nicht unbedingt in Abhängigkeit zur Verwendung von Laiensprache oder leicht verständlicher Sprache.

Die Analysen der Inter-Rater-Reliabilität zeigten, dass sich der Grad der Übereinstimmung in den Bewertungen aller drei Kategorien zwischen beiden Prüfergruppen teilweise unterschied. Die Prüfenden der „Was hab' ich?“-GmbH hatten in der Kategorie „Inhaltsauswahl und medizinische Korrektheit“ einen höheren Grad an Übereinstimmung als die Prüfenden des IMPP und die Allgemeinmediziner*innen. Diese Tatsache ist ein Indiz für unterschiedliche Ausgangsbedingungen der Prüfenden. Dies ist auf ihre interdisziplinären Hintergründe und vor allem auf ihre unterschiedlichen Vorerfahrungen mit patientenverständlicher Sprache zurückzuführen. Dieses Ergebnis unterstreicht die Notwendigkeit einer einheitlichen Prüferschulung zu den Standards des Verfassens patientenverständlicher Berichte, bevor der Bewertungsbogen in den medizinischen Staatsexamina verwendet wird.

Der berechnete G-Koeffizient zeigt, dass die Daten und Ergebnisse der Studie auf weitere Bewertungen außerhalb der Studie anwendbar und daher verallgemeinerbar sind. Der Bewertungsbogen generiert somit reliable Ergebnisse. Mit der Sicherstellung der inhaltlichen Validität des Bewertungsbogens durch den Entwicklungsprozess wurden die Testgütekriterien weitgehend erfüllt. Der Bogen ist somit ein geeignetes Instrument, um patientenverständliche Berichte mittels zentraler Kategorien zu bewerten.

Zu den Einschränkungen der Studie gehört, dass die Ergebnisse nur aus Berichten stammten, die auf einer interprofessionellen Ausbildungsstation in Heidelberg verfasst wurden. Im Rahmen der weiteren Forschung wäre es interessant, die Bewertungen mehrerer Prüfender in einer größeren Stichprobe von Berichten, die auf konventionellen Stationen oder im ambulanten Bereich verschiedener Fakultäten verfasst wurden, zu evaluieren und zu validieren. Zudem sollten Prüferschulungen zur Beurteilung von patientenverständlichen Berichten unter Verwendung des entwickelten Bewertungsbogens obligatorisch sein. Diese Schulungen sollen einen einheitlichen Bewertungsstandard gewährleisten, um unterschiedliche Interpretationen und Gewichtungen von Unterpunkten zu vermeiden und zu einer höheren Beurteilungsübereinstimmung beizutragen. Die Inter-Rater-Reliabilität der geschulerten Prüfenden sollte anschließend analysiert werden. Kontinuierliche formative Prüfungen und Feedback auf der Grundlage des Bewertungsbogens können eingesetzt werden, um die medizinische Ausbildung für PJ-Studierende zu verbessern. Da Patient*innen die Empfänger*innen dieser Berichte sind, wäre ein Vergleich der Beurteilungen der patientenverständlichen Berichte durch die Patient*innen selbst mit der Beurteilung durch medizinische Expert*innen von großem Interesse. Durch Kombination beider Beurteilungen könnte sichergestellt werden, dass die patientenverständlichen Berichte Patient*innen umfassend informieren und Entlassungsfehler in Zukunft reduziert werden können. Die Verbesserung der mündlichen Kommunikationsfähigkeiten von Studierenden und jungen Ärzt*innen aufgrund dieser Weiterentwicklung der medizinischen Ausbildung sollte in weiteren Studien untersucht werden.

5. Schlussfolgerung

Die Bewertung der Kommunikation mit Patient*innen ist ein Zugewinn der neu entwickelten Prüfung am Patienten oder an der Patientin. Regelmäßige formative Prüfungen und Feedback anhand des Bewertungsbogens können im Laufe des Studiums eingesetzt werden, um die schriftlichen Kommunikationskompetenzen künftiger Ärzt*innen zu verbessern. Darüber hinaus ist zu erwarten, dass das Verstehen von Diagnosen und Therapien sowie eine vertrauensvolle Arzt-Patienten-Beziehung Patient*innen zu mehr Teilhabe am medizinischen Entscheidungsprozess befähigt. Dies könnte künftig auch Entlassungs-

fehlern vorbeugen. Um dieses Ziel zu erreichen sind klare Vorgaben sowie Trainings zum Verfassen patientenverständlicher Berichte in die medizinische Ausbildung zu verankern. Für den einheitlichen Gebrauch des Bewertungsbogens ist die Implementierung standardisierter Prüferschulungen erforderlich.

Abkürzungsverzeichnis

- G-Koeffizient: Generalisierbarkeitskoeffizient
- HIPSTA: Heidelberger Interprofessionelle Ausbildungsstation
- IMPP: Institut für medizinische und pharmazeutische Prüfungsfragen
- MME: Master of Medical Education
- PJ: Praktisches Jahr
- r: Pearsons Korrelationskoeffizient r

Förderung

Dieses Projekt wurde von der Bertelsmann Stiftung gefördert (Laufzeit: 1.10.2017 - 30.06.2021).

Aktuelle berufliche Rolle der Autor*innen

Dr. med. Lena Selgert

- wissenschaftliche Mitarbeiterin am Institut für medizinische und pharmazeutische Prüfungsfragen, Mainz
- Ärztin

Bernd Bender

- wissenschaftlicher Mitarbeiter am Institut für medizinische und pharmazeutische Prüfungsfragen, Mainz
- Diplom-Soziologe

Dr. phil. Barbara Hinding

- wissenschaftliche Mitarbeiterin am Institut für medizinische und pharmazeutische Prüfungsfragen, Mainz
- Diplom-Psychologin

Aline Federmann, M.A.

- wissenschaftliche Mitarbeiterin am Institut für medizinische und pharmazeutische Prüfungsfragen, Mainz

Prof. Dr. med. André L. Mihaljevic

- Oberarzt und Lehrkoordinator klinischer Studienabschnitt (Klinik für Allgemein-, Viszeral- und Transplantationschirurgie, Universitätsklinikum Heidelberg)
- Ärztliche Leitung der Heidelberger Interprofessionellen Ausbildungsstation – (HIPSTA)
- Klinischer Wissenschaftler (stellv. Sprecher der chirurgischen Studiennetzwerks CHIR-Net)

Rebekka Post

- Ärztin bei der Initiative "Was hab' ich" gGmbH

Ansgar Jonietz

- Gründer und CEO der Initiative "Was hab' ich? gGmbH
- Informatiker

John J. Norcini, Ph.D.

- Research Professor in the Department of Psychiatry at SUNY Upstate Medical University
- President Emeritus of FAIMER

Ara Tekian, Ph.D., MHPE

- professor, Department of Medical Education, and associate dean, International Education, University of Illinois at Chicago College of Medicine, Illinois, USA.

Prof. Dr. med. Jana Jünger, MME (Bern)

- Direktorin des Institutes für medizinische und pharmazeutische Prüfungsfragen
- Entwicklung des Postgraduierten-Studiengangs Master of Medical Education (MME), Deutschland
- Mitglied der MME-Studiengangsleitung und Dozentin der Module Prüfungen, Ausbildungsforschung, Evaluation
- Leitung verschiedener Projekte zur Implementierung von Kommunikationscurricula in die medizinische Ausbildung sowie Entwicklung neuer Prüfungsformate zur Überprüfung kommunikativer Kompetenzen

Interessenkonflikt

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

Literatur

1. Bundesministerium für Gesundheit. Patientenrechte. Berlin: Bundesministerium für Gesundheit; 2019. Zugänglich unter/available from: <https://www.bundesgesundheitsministerium.de/themen/praevention/patientenrechte/patientenrechte.html>
2. Stahl K, Nadj-Kittler M. Picker report 2016. Vertrauen braucht gute Verständigung. Erfolgreiche Kommunikation mit Kindern, Eltern und erwachsenen Patienten. Hamburg: Picker Institut Deutschland gGmbH; 2016.
3. NHS England. Review of National Reporting and Learning System (NRLS) incident data relating to discharge from acute and mental health trusts - August 2014. London: NHS England; 2014.
4. Williams H, Edwards A, Hibbert P, Rees P, Prosser Evans H, Panesar S, Carter B, Parry G, Makeham M, Jonas A, Avery A, Sheikh A, Donaldson L, Carson-Stevens A. Harms from discharge to primary care: mixed methods analysis of incident reports. *Br J Gen Pract.* 2015;65(641):829-837. DOI: 10.3399/bjgp15X687877
5. Hesselink G, Zegers M, Vernooij-Dassen M, Barach P, Kalkman C, Flink M, Öhlen G, Olsson M, Bergenbrant S, Orrego C, Suñol R, Toccafondi G, Venneri F, Dudzik-Urbaniak E, Kutryba B, Schoonhoven L, Wollersheim H; European HANDOVER Research Collaborative. Improving patient discharge and reducing hospital readmissions by using Intervention mapping. *BMC Health Serv Res.* 2014;14:389. DOI: 10.1186/1472-6963-14-389

6. Pinelli V, Papp KK, Gonzalo JD. Interprofessional communication patterns during patient discharges: A social network analysis. *J. Gen Intern Med.* 2015;30(9):1299-306. DOI: 10.1007/s11606-015-3415-2
7. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in Communication and Information Transfer between Hospital-Based and Primary Care Physicians: Implications for Patient Safety and Continuity of Care. *JAMA.* 2007;297(8):831-841. DOI: 10.1001/jama.297.8.831
8. Noordman J, van Vliet L, Kaunang M, van den Muijsenbergh M, Boland G, van Dulmen S. Towards appropriate information provision for and decision-making with patients with limited health literacy in hospital-based palliative care in Western countries: a scoping review into available communication strategies and tools for healthcare providers. *BMC Palliat Care.* 2019;18(1):37. DOI: 10.1186/s12904-019-0421-x
9. Jünger J, Mutschler A, Kröll K, Weiss C, Fellmer-Drüg E, Köllner V, Ringel N. Ärztliche Gesprächsführung in der medizinischen Aus- und Weiterbildung: Das nationale longitudinale Mustercurriculum Kommunikation. *Med Welt.* 2015;66: 189-192.
10. Sator M, Jünger J. From Stand-Alone Solution to Longitudinal Communication Curriculum -Development and Implementation at the Faculty of Medicine in Heidelberg. *Psychother Psych Med.* 2015;65(05):191-198. DOI: 10.1055/s-0034-1398613
11. Jünger J, Weiss C, Fellmer-Drüg E, Semrau J. Verbesserung der kommunikativen Kompetenzen im Arztberuf am Beispiel der Onkologie: Ein Projekt des Nationalen Krebsplans. *Forum.* 2016;31:473-478. DOI: 10.1007/s12312-016-0162.1
12. van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205-214. DOI: 10.3109/0142159X.2012.652239
13. Nikendei C, Jünger J. OSCE - hands on instructions for the implementation of an objective structured clinical examination. *GMS Z Med Ausbild.* 2006;23(3):Doc47. Zugänglich unter/available from: <https://www.egms.de/static/de/journals/zma/2006-23/zma000266.shtml>
14. Jünger J. Kompetenzorientiert prüfen im Staatsexamen Medizin. *Bundesgesundheitsbl.* 2018;61:171-177. DOI: 10.1007/s00103-017-2668-9
15. Schwartzberg JG, Cowett A, VanGeest J, Wolf MS. Communication techniques for patients with low health literacy: a survey of physicians, nurses, and pharmacists. *Am J Health Behav.* 2007;Suppl 1:96-104. DOI: 10.5555/ajhb.2007.31.supp.S96
16. Earl GL, Harris EM, Dave M, Estriplet-Jiang, J. Implementing a health literacy module fostering patient-centered written communication in a cardiovascular prevention elective course. *Curr Pharm Teach Learn.* 2019;11(7):702-709. DOI: 10.1016/j.cptl.2019.03.008
17. Lopez Ramos C, Williams JE, Bababekov YJ, Chang DC, Carter BS, Jones PS. Assessing the Understandability and Actionability of Online Neurosurgical Patient Education Materials. *World Neurosurg.* 2019;130:588-597. DOI: 10.1016/j.wneu.2019.06.166
18. Roberts HJ, Zhang D, Earp BE, Blazar P, Dyer GSM. Patient self-reported utility of hand surgery online patient education materials. *Musculoskeletal Care.* 2018;16(4):458-462. DOI: 10.1002/msc.1360
19. Davis TC, Fredrickson DD, Arnold C, Murphy PW, Herbst M, Bocchini JA. A polio immunization pamphlet with increased appeal and simplified language does not improve comprehension to an acceptable level. *Patient Educ Nurs.* 1998;33(1):25-37. DOI: 10.1016/s0738-3991(97)00053-0
20. Rubin DT, Ulitsky A, Poston J, Day R, Huo D. What is the most effective way to communicate results after endoscopy? *Gastrointest Endosc.* 2007;66(1):108-112. DOI: 10.1016/j.gie.2006.12.056
21. Schumaier AP, Kakazu R, Minoughan CE, Grawe BM. Readability assessment of American Shoulder and Elbow Surgeons patient brochures with suggestions for improvement. *JSES Open Access.* 2018;2(2):150-154. DOI: 10.1016/j.jses.2018.02.003
22. "Was hab' ich?" gGmbH. Patientenbriefe wirken. Ergebnisbericht zum Projekt "Mehr Gesundheitskompetenz durch Patientenbriefe". Hamburg: "Was hab' ich?" gGmbH; 2019.
23. Netzwerk Leichte Sprache. Die Regeln für Leichte Sprache. Münster: Netzwerk Leichte Sprache; 2017. Zugänglich unter/available from: https://www.leichte-sprache.org/wp-content/uploads/2017/11/Regeln_Leichte_Sprache.pdf
24. Bredele U, Maaß C. Duden Leichte Sprache. Theoretische Grundlagen, Orientierung für die Praxis. Berlin: Dudenverlag; 2016.
25. Selgert L, Samigullin A, Lux R, Gornostayeva M, Hinding B, Schlasius-Ratter U, Hendelmeier M, Mihaljevic AL, Wienand S, Schneidewind S, Bintaro P, Jonitz A, Jünger J. Weiterentwicklung des medizinischen Staatsexamens in Deutschland: Prüfung am Patienten. In: Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA), des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ) und der Chirurgischen Arbeitsgemeinschaft Lehre (CAL). Frankfurt am Main, 25.-28.09.2019. Düsseldorf: German Medical Science GMS Publishing House; 2019. DocP-05-03.DOI: 10.3205/19gma287
26. Mihaljevic AL, Schmidt J, Mitzkat A, Probst P, Kenngott T, Mink J, Fink CA, Ballhausen A, Chen J, Cetin A, Murrmann L, Müller G, Mahler C, Götsch B, Trierweiler-Hauke B. Heidelberger interprofessionelle Ausbildungsstation (HIPSTA): a practice- and theory-guided approach to development and implementation of Germany's first interprofessional training ward. *GMS J Med Educ.* 2018;35(3):Doc33. DOI: 10.3205/zma001179
27. Post R, Jonietz A, Selgert L, Lux R, Mihaljevic AL, Jünger J. Entwicklung, Testung und Validierung eines Bewertungsbogens zur Beurteilung laienverständlicher Patientenbriefe. In: Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA), des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ) und der Chirurgischen Arbeitsgemeinschaft Lehre (CAL). Frankfurt am Main, 25.-28.09.2019. Düsseldorf: German Medical Science GMS Publishing House; 2019. DocP-05-02. DOI: 10.3205/19gma286
28. Weins C. Uni- und bivariate deskriptive Statistik. In: Wolf C, Best H, editors. *Handbuch der sozialwissenschaftlichen Datenanalyse.* Wiesbaden: VS-Verlag für Sozialwissenschaften; 2010. p.65-89. DOI: 10.1007/978-3-531-92038-2_4
29. Field S, Egan R, Beesley T. Applying G-Theory and Multivariate G-Analysis to improve Clinical Data quality and performance assessment accuracy. OHSE Working Paper. San Francisco, CA: Academia.edu; 2017. Zugänglich unter/available from: https://www.academia.edu/37118667/_Applying_G-Theory_and_Multivariate_G-Analysis_to_improve_Clinical_Data_quality_and_performance_assessment_accuracy_2017_OHSE_Working_Paper
30. Schmolck P. Begleittext: Methoden der Reliabilitätsschätzung. München: Universität der Bundeswehr München; 2007. Zugänglich unter/available from: <https://dokumente.unibw.de/pub/bscw.cgi/1787978>
31. Cardinet J, Tourneur Y, Allal L. Extension of generalizability theory and its applications in educational measurement. *J Educ Measurement.* 1981;18(4):183-204. DOI: 10.1111/j.1745-3984.1981.tb00852.x
32. Bühl A. SPSS 16. Einführung in die moderne Datenanalyse. 11th ed. München: Pearson-Studium; 2008.

33. Döring N, Bortz J. Datenerhebung. In: Döring N, Bortz J, editors. Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Berlin, Heidelberg: Springer; 2016. p.321-577. DOI: 10.1007/978-3-642-41089-5_10
34. Bortz J, Lienert GA, Boehnke K. Verteilungsfreie Methoden in der Biostatistik. 3rd ed. Heidelberg: Springer Medizin Verlag; 2008.
35. Rammstedt B. Reliabilität, Validität, Objektivität. In: Wolf C, Best H, editors. Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS-Verlag für Sozialwissenschaften; 2010. p.239-258. DOI: 10.1007/978-3-531-92038-2_11
36. Bittner A, Bittner J, Jonietz A, Dybowski C, Harendza S. Translating medical documents improves students' communication skills in simulated physician-patient encounters. BMC Med Educ. 2016;16:72. DOI: 10.1186/s12909-016-0594-4
37. Operation Karriere. Unsere Vision ist es, Arzt und Patient auf Augenhöhe zu bringen. Köln: Deutscher Ärzteverlag GmbH; 2017. Zugänglich unter/available from: <https://www.operation-karriere.de/karriereweg/bewerbung-berufsstart/unsere-vision-ist-es-arzt-und-patient-auf-augenhoehe-zu-bringen.html>
38. Kühnel SM, Krebs D. Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. 4th ed. Reinbek: Rowohlt; 2007.
39. Schmidt RC. Managing Delphi surveys using nonparametric statistical techniques. Decision Sci. 1997;28(3):763-774. DOI: 10.1111/j.1540-5915.1997.tb01330.x

Korrespondenzadresse:

Dr. med. Lena Selgert
 Institut für medizinische und pharmazeutische
 Prüfungsfragen (IMPP), Postfach 2518, 55015 Mainz,
 Deutschland
 LSelgert@impp.de

Bitte zitieren als

Selgert L, Bender B, Hinding B, Federmann A, Mihaljevic AL, Post R, Jonietz A, Norcini J, Tekian A, Jünger J. Development, testing and generalizability of a standardized evaluation form for the assessment of patient-directed reports in the new final medical licensing examination in Germany. GMS J Med Educ. 2021;38(3):Doc71. DOI: 10.3205/zma001467, URN: urn:nbn:de:0183-zma0014672

Artikel online frei zugänglich unter

<https://www.egms.de/en/journals/zma/2021-38/zma001467.shtml>

Eingereicht: 31.03.2020

Überarbeitet: 10.08.2020

Angenommen: 21.09.2020

Veröffentlicht: 15.03.2021

Copyright

©2021 Selgert 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/>.