UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

The general practice specialist examination - examiners' perspectives on the status quo and future development: an exploratory qualitative study = Die Facharztprüfung Allgemeinmedizin –Perspektiven von Prüfenden auf Status Quo und Weiterentwicklung: eine explorative, qualitative Studie

Andreas C. Dreher, Christine Becker, Marco Roos, Attila Altiner, Simon Schwill

Angaben zur Veröffentlichung / Publication details:

Dreher, Andreas C., Christine Becker, Marco Roos, Attila Altiner, and Simon Schwill. 2026. "The general practice specialist examination - examiners'perspectives on the status quo and future development: an exploratory qualitative study = Die Facharztprüfung Allgemeinmedizin – Perspektiven von Prüfenden auf Status Quo und Weiterentwicklung: eine explorative, qualitative Studie." *GMS Journal for Medical Education* 43 (6): Doc78. https://doi.org/10.3205/zma001872.

Nutzungsbedingungen / Terms of use: **CC BY 4.0**

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
CC-BY 4.0: Creative Commons: Namensnennung
Weitere Informationen finden Sie unter: / For more information see:
https://creativecommons.org/licenses/by/4.0/deed.de

The general practice specialist examination – examiners' perspectives on the status quo and future development: An exploratory qualitative study

Andreas Christian Dreher[1]
Christine Becker[1]
Marco Roos[2]
Attila Altiner[1]
Simon Schwill[1]

1 University Hospital Heidelberg, Department for Primary Care and Health Services Research, Heidelberg, Germany

2 University of Augsburg, Faculty of Medicine, Department of General Practice, Augsburg, Germany

Abstract

Background: The specialist examination has a substantial impact on the quality of healthcare delivery. The aim of this study was to explore examiners' perspectives on the current examination format, their motivation for serving as examiners, perceived strengths and weaknesses of the existing assessment, suggestions for improvement, and the perceived effects of Competence Centres for Postgraduate Training (CCPTs).
Methods: Examiners from across Germany were recruited and interviewed in semi-structured video interviews using a self-developed interview guide. Interviews were transcribed verbatim and analysed using qualitative content analysis.
Results: Twelve examiners from six federal states participated. The mean interview duration was 33 minutes (range: 24-57 minutes). Key advantages identified included the one-to-one examination setting, the ability to ensure a minimum standard of specialist competence and the examination's efficiency in terms of time and resources. Oral examinations were perceived as having limited ability to assess the complex clinical competencies required in patient care. Perceived needs for change varied among participants. Suggested improvements included examiner training, the development of shared assessment criteria and the introduction of practical examination components. Longitudinal assessments were also mentioned.
Discussion: Examiners recognised limitations of the current examination regarding objectivity, reliability, validity and contextual sensitivity, particularly in relation to the complexity of general practice. Nevertheless, they considered the assessment of a minimum standard of specialist competence to be feasible. The findings reveal a tension between practical feasibility and educational assessment standards. Potential areas for improvement include examiner training, regular exchange among examiners, shared assessment frameworks, a clearer definition of examination objectives and the implementation of longitudinal assessment approaches (e.g. portfolio-based assessment with regular interim assessments).
Conclusions: The specialist examination demonstrates considerable potential for further development. Shared assessment frameworks, as developed during examiner training, could provide greater structure to the examination process and may be implemented with relative ease.

Keywords: specialist examination, examiner training, postgraduate medical education, Competence Centres for Postgraduate Training, general practice

Introduction

The specialist examination in general practice marks the completion of five years of full-time equivalent postgraduate training [1]. In Germany, the German medical association is responsible for administering these examinations, which serve as the formal certification of specialist competence. As the final assessment in a physician's professional training pathway, successful completion permits independent and autonomous (specialist) practice. Consequently, the specialist examination represents an important instrument for ensuring specialist qualification, with direct implications for healthcare quality and patient safety [2], [3]. The examination usually takes the form of an oral assessment lasting at least 30 minutes and is conducted by two specialist examiners and a chairperson. Traditionally, it consists of an unstructured collegial oral discussion, a format that has been firmly established in Germany for decades. However, its implementation may vary according to region, specialty and individual examination practices.

The introduction of a five-year postgraduate training programme and a specialist examination in general practice represented a major step towards achieving parity between general practice and other medical specialties. In recent years, the number of specialist examinations in general practice has increased considerably [4]. In Baden-Württemberg (federal state of Germany), for example, the examination currently constitutes the most frequently administered specialist examination [5]. Failure rates in German specialist examinations are generally in the single-digit percentage range, whereas the European specialist examination in emergency medicine reports failure rates of up to 64% [6]. In addition to workplace-based training under supervision, Competence Centres for Postgraduate Training (CCPTs) have been established in Germany to enhance the quality and efficiency of postgraduate training in general practice [7]. These centres support trainees through structured seminar programmes, mentoring and participation in regional training networks.

In general, assessments should be objective, valid, reliable [8], [9] and sensitive to the context in which professional performance occurs [3]. Within medical education, assessment is considered a key driver of learning ("assessment drives learning") [10], [11]. Furthermore, competency-based, practical and standardised assessment formats have gained increasing importance over recent decades [12], [13] and have become well established at many universities. Concepts such as Competency-Based Medical Education (CBME) and Entrustable Professional Activities (EPAs) have moved increasingly into focus [14]. The revised German Model Specialty Training Regulations [1] have likewise adopted a competency-based orientation; however, the format of the specialist examination has remained largely unchanged. From a psychometric perspective, the current specialist examination is insufficient for the valid assessment of competency-based outcomes and practical clinical skills [15]. According to the principle of constructive alignment [16], learning objectives, educational activities and assessment methods should be closely aligned to facilitate goal-directed learning. At present, alignment between the competency objectives of the MWBO and the predominantly unstructured oral examination format appears limited. From an educational assessment perspective, substantial further development of the specialist examination is therefore warranted to achieve better alignment with the competency goals of the revised MWBO. As the final component of postgraduate training in general practice, the specialist examination has important implications for healthcare quality. A recent survey among general practice trainees and specialists found little demand for changes to the current examination format [17]. The aim of this exploratory study was therefore to investigate how examiners evaluate the current examination format, what motivates them to undertake examining activities, which strengths and weaknesses they attribute to the existing assessment, which areas for further development they identify, what qualifications they consider necessary for examiners and to what extent they perceive the Competence Centres for Postgraduate Training to influence examination performance.

Material and methods

Study design and setting

A qualitative study using semi-structured interviews with examiners in general practice was conducted. A qualitative approach was chosen to enable an in-depth exploration of experiences and perceived challenges in their complexity and contextual embeddedness. Recruitment, interviews and subsequent data analysis were carried out over a 16-week period between December 2024 and April 2025. Reporting of the study followed the *Consolidated Criteria for Reporting Qualitative Research* (COREQ) guidelines (see attachment 1) [18].

Participants and recruitment

All participants were examiners in general practice and were contacted via email. As no publicly available national register or network of specialist examiners exists in Germany, purposive sampling was employed to ensure variation with regard to federal state, examination experience and gender.

Examiners were recruited both through existing professional networks (Competence centres for postgraduate training and professional conferences) and through the targeted identification of previously unknown individuals based on examination records. Some participants were known to the researchers through professional contacts. However, no hierarchical or other dependent relationship existed between researchers and participants.

A total of 40 examiners were contacted, of whom 13 responded and 12 ultimately participated in an interview.

Development of the interview guide

Following a literature review (NIH/PubMed® search terms: *"specialist exam", "board certification", "resident assessment"* and *"postgraduate medical education assessment"*), an interview guide was developed by a team of four researchers (ACD, CB, MR and SS) with expertise in qualitative research, general practice and medical education.

Kern's six-step approach to curriculum development [19] served as the theoretical framework for structuring the study, as it systematically links needs assessment, goal definition, educational activities and assessment. Both a needs assessment of the target group of examiners and the positioning of the specialist examination within the broader context of postgraduate training were considered during the development of the interview guide.

A semi-structured interview format was chosen to ensure systematic coverage of the key thematic domains while simultaneously allowing for free associations and follow-up questions.

The interview guide was validated using the think-aloud technique [20], in which participants verbalise their cognitive processes, strategies and decision-making while responding to questions. Subsequently, the interview guide was piloted in two interviews and revised slightly (see attachment 2). The final interview guide included questions on sociodemographic characteristics, experience as an examiner, perceptions of the current specialist examination, perspectives on potential changes to the examination format and possible effects of the Competence Centres for Postgraduate Training.

Data collection

Interviews were conducted via Zoom® (San Jose, California, USA; version 6.3.0) by ACD (specialist in psychosomatic medicine and psychotherapy, physician in advanced postgraduate training in general practice, examiner for psychological psychotherapists and child and youth psychotherapists, with several years of experience in conducting interviews and qualitative data analysis).

At the beginning of each interview, participants were informed about the background of the study and the role of the researcher. No third parties (non-participants) were present either at the researcher's workplace or in the private environments of the participants during the interviews. Audio recordings were made for all interviews. Field notes were taken to document relevant observations made during data collection. Interview transcripts were not returned to participants for commenting or correction.

Qualitative data analysis

All interviews were transcribed verbatim and analysed manually using a paper-based approach. Transcripts were reviewed and coded using an inductive-deductive qualitative content analysis approach based on Kuckartz [21]. Two researchers (ACD and SS) independently analysed the material. Categories and representative quotations were subsequently compared and revised where necessary. In cases of disagreement or uncertainty, a third researcher (CB) reviewed the material in order to reach consensus.

Throughout the analytical process, continuous reflexivity was maintained regarding personal, interpersonal, methodological and contextual factors that might have influenced data interpretation.

Results

Interview characteristics

A total of 12 interviews were conducted between December 2024 and April 2025. The mean interview duration was 33 minutes (range: 24-57 minutes). No technical problems occurred during data collection and none of the interviews had to be repeated.

The coding framework was reviewed after the first two interviews and no further modifications were required. No new themes or codes emerged in the final interviews, indicating that meaning saturation had been achieved and no additional participants were recruited.

Participant characteristics and examination experience

Participants' sociodemographic characteristics are presented in table 1. All participating examiners were specialists in general practice and possessed substantial expertise in conducting specialist examinations. Across the participating federal states, differences existed regarding whether the chairperson of the examination board was required to belong to the specialty of general practice or another medical discipline. Furthermore, audio recording of examinations was not implemented consistently across federal states.

Motivation for serving as an examiner

Motivation for undertaking examiner responsibilities were predominantly driven by idealistic and intrinsic factors (see table 2). Financial compensation, which was generally perceived as insufficient to offset lost income, was identified as a potential barrier to recruiting new examiners.

Strengths and weaknesses of the current examination format

Perceived strengths and weaknesses of the current examination format are summarised in table 3. Representative quotations are presented in attachment 3.

Most examiners identified organisational strengths as well as content-related advantages of the current examination format. A major strength reported by the majority

Table 1: Sociodemographic characteristics of participating specialist examiners in general practice (N=12)

Age	**In years, median (IQR)**	**62 (56, 66)**
Gender	Female	7
	Male	5
Federal state	Baden-Württemberg Bavaria Bremen Hesse Lower Saxony Saxony	3 2 1 2 1 3
Specialist qualification	General Practice	12
experience as examiner	Years, mean (SD) Years, median (IQR)	13,42 (9,91) 12 (5,20)
Number of examinations conducted	Number of examinations conducted, mean (SD) Number of examinations conducted, median (IQR)	200,67 (82,5) 178 (136,254)

SD=standard deviation; IQR=interquartile range

Table 2: Reasons for serving as a specialist board examiner

Motivation related to the discipline	Quality assurance	*#1 "I consider it an important activity in the context of quality assurance in general practice."*
	Contribution to postgraduate training	*#2 "My motivation is the wish to contribute to postgraduate training and for me this also includes taking on the burden of conducting examinations."*
Intrinsic/personal motivation	Honour/professional commitment	*#7 "Personally, it is an honour for me to conduct specialist examinations."* *#8: "So, out of the conviction that someone has to do it"*
	Remaining professionally up to date	*#9 "It forces me to keep my medical knowledge up to date."*
Societal and structural motives	Representation of the discipline	*#10 "General practitioners should examine general practitioners."*
	Gender representation	*#10 "As the medical profession becomes increasingly female, we also need female examiners."*

of participants was the feasibility of conducting the examination with relatively little time investment and good practical applicability. Examinations conducted by specialist peers were considered to provide professionally sound quality assurance and to enable the assessment of professional attitudes. Some participants additionally regarded the individual face-to-face examination setting and the flexibility in selecting questions and topics as advantages.

At the same time, several limitations were identified. Most participants criticised the limited objectivity of oral examinations. In particular, the validity of the oral format for assessing practical patient care was questioned, especially in relation to the complexity of long-term care for multimorbid patients in general practice.

> *"We only see the idealised picture that the candidate verbally presents to us about how they would manage the cases discussed. We have no opportunity to observe their everyday practice, their actual performance or their doctor–patient communication." (#1)*

Some participants further identified structural disadvantages, including the late timing of the examination at the end of training, the limited opportunity to influence the educational process, examiner uncertainty and the administrative burden associated with failed examinations. Within the discussion of strengths and weaknesses, participants emphasised that competencies not directly assessed during the examination may nevertheless be present and that the examination is only of limited suitability for comprehensively assessing competencies within the available timeframe.

> *"In my view, it is impossible to pursue too many objectives within 30 minutes. The purpose is for examiners to determine whether the candidate is capable of working independently and unsupervised as a general practitioner in their own practice." (#2)*

Table 3: Advantages and disadvantages of the current specialist examination format as perceived by examiners

Advantages		Disadvantages	
Organisational advantages	Accessibility	Lack of objectivity and standardisation	Subjectivity (examiner dependence, differing standards)
	Practical feasibility		Lack of standardised conduct
	Low time requirement		Desire for defined assessment criteria (marking scheme)
Professional quality	Assessment by specialist peers	Limited validity	Limited relevance to actual clinical practice
	Quality assurance		Limited ability to capture the complexity of general practice
	Assessment of professional attitudes		Difficulty distinguishing performance within a short time
Advantages of the examination setting	Individual examination	Structural disadvantages	Examination only at the end of postgraduate training
	Personal interaction		Burden and uncertainty for examiners when candidates fail
Flexibility for examiners	Freedom to choose questions and topics	Potential for manipulation or exploitation	Risk of candidates exploiting the examination format

Criteria for becoming an examiner

All participants reported that they had been appointed by their regional medical association and that, to their knowledge, no transparent selection process existed. Table 4 summarises the criteria participants considered important for the selection of examiners.

Several participants questioned whether objective criteria alone would adequately capture personal suitability for the examiner role. One examiner proposed collegial discussions based on case analyses as a potential approach for assessing both professional and personal competencies of prospective examiners.

Perspectives on future developments

Participants expressed differing views regarding the need for changes to the current examination format (see table 5). Some considered the status quo satisfactory and opposed what they described as "schoolification" of specialist examinations according to university assessment principles.

The majority, however, favoured optimisation and further development of the existing format. In particular, participants advocated the introduction of shared assessment frameworks and predefined expectations in order to improve the examination process. Some additionally proposed supplementing the current examination with written and practical assessment components.

Three participants expressed support for longitudinal assessment models based on international examples, arguing that such approaches would better capture developmental processes and competencies over time. One participant suggested that participation in such assessments could be voluntary.

Perceptions of the potential impact of changes to the examination on postgraduate training varied. While some participants anticipated little effect, others considered examination reform to offer considerable opportunities for improving training quality.

> *"In my opinion, changing the examination could make a substantial contribution to improving both the quality of and satisfaction with postgraduate training. If assessment became broader, trainees would have to adapt, learn differently and place different demands on their trainers." (#10)*

Overall, most participants agreed that the primary emphasis should be placed on improving the quality of postgraduate training rather than creating increasingly difficult examinations.

Perceived effects of the competence centres for postgraduate training

Approximately half of the interviewed examiners were aware whether candidates had participated in a Competence Centre for Postgraduate Training, for example through active enquiry or prior acquaintance from seminars and were therefore able to comment on possible effects.

These participants reported that trainees who had attended a CCPT appeared to perform better in examinations and demonstrated greater methodological competence and more structured approaches in answering questions.

Table 4: Criteria identified for examiners in general practice from the examiners' perspective

Professional competence and work experience	Relevant experience in general practice work	#5: "Yes, work as a GP, but I think that is an implicit criterion. I can only assess well and fairly if I have actually been working in general practice for some time."
	Evidence-based, guideline-oriented practice	#10: "So what is very important is evidence-based medicine, whether people work according to guidelines or also know when they can deviate from guidelines."
	Reflective general practice approach	#9: "General practice is more than just the evidence from studying. Can someone reflect on their actions in relation to general practice work and also critically question themselves?"
Examination and teaching experience	Prior experience in designing and conducting examinations	#6: "So examination experience in other contexts. There are exams, for example people involved in medical assistant (MFA - Medizinische Fachangestellte) examinations, who already have some understanding of what happens in examinations and how to deal with candidates."
	General teaching experience	#11: "Secondly, I would find it good if people had some teaching experience."
	Postgraduate training experience/ authorisation	#9: "A training licence, definitely, that they have also trained others. Perhaps not only the formal authorisation but also actual experience in postgraduate training."
Examiner training / shared standards	Participation in examiner training	#2: "And that afterwards I would conduct examiner training, where we develop ideas together, let them develop case scenarios and have them conduct an examination. Ideally even with video and audio recordings and then discuss whether there are areas for improvement."
	Examiners being examined ("examining the examiners")	#2: "I would first suggest that every examiner must undergo an examination themselves. I would find it very interesting if examiners were assessed through a collegial conversation. It does not need to be a formal exam setting, but that you sort of 'embed' them and speak with them for an hour: how they organise their working life, patient communication, and case management. Let's go through a few cases together."
	Obligation to create expected standards (marking schemes)	#11: "That one is obliged to create and discuss an expected standard beforehand. That is about two pages of A4, but people should be willing to do that."
Personal suitability and soft skills	Teamwork ability	#10: "Then what is important to me is a team player. People who are also able to function well as part of a team."
	Personal integrity	#9: "And personal suitability: personalities are difficult to categorise. One should also be kind, balanced and empathetic. But it is difficult to define clear criteria for personal suitability."
	Willingness to conduct examinations	#1: "It is difficult enough to find people who are willing to take on the time commitment and the loss of clinic time." #7: "The number one priority is the willingness to do it."
Diversity in examination committees	Socio-demographic characteristics	#9: "Then I actually use a mix of who we would like to have: women and men, younger and older in the composition of the committees."
	Backgrounds of experience	#4: "This is something very important, which is actually not represented: how to balance economic decision-making, staff management, guidelines and patient care. An examiner should reflect this in their personal background and practical experience."
	Self-employment/ employment status	#3: "In our current system, being a self-employed GP is important in order to assess answers that may differ from one's own perspective. This includes economic considerations and workflow considerations."

They also attributed higher levels of practical clinical competence to these trainees.

At the same time, participants acknowledged that distinguishing the specific effects of CCPT participation from other influencing factors was difficult. For example, diverse training experiences involving several ambulatory rotations were considered an important predictor of examination performance. Examiners also emphasised that non-participants could perform equally well in specialist examinations.

Nevertheless, a commonly perceived trend was that most candidates who failed the specialist examination had not participated in a CCPT programme.

Table 5: Specialist examiners' perspectives on changes to the current examination format

Preservation of the status quo	No change, satisfaction with the status quo	#4: "I honestly find the idea of introducing change with regular assessments bureaucratic and frankly stupid, because everyone has to design their own residency anyway. And in the end, general practice is as diverse as society needs different doctors and types. I think the way it is now is good."
	No "schooling" of the examination using rigid assessment criteria	#4: "No, I would be very against making it even more like school"
Improvement of examination organisation	Exchange between examiners before the examination	#8: "A brief exchange between the two examiners two or three days beforehand. As it is, the candidates only find out on the day of the exam who will examine them. We already know that, but we are also not told with whom we will be examining. I would find it good if we could exchange information beforehand."
	Desire for internal criteria for the examination/expected standards	#9: "So I would like to have some internal criteria defining what should definitely be examined, which areas, and what a basic passing standard would be, and what goes beyond that. So you have candidates where you say: at least he hasn't killed anyone. And then there are these people where you think: hey, give me your address, I want to hire you immediately. So some basic criteria would be enough for passing."
Further development of the existing examination format	Current examination format with diversified methods	#10: "The use of more real materials, such as ECGs, images, ultrasound scans, examination techniques whether with phantoms or simulated patients. I would like that. And of course video, meaning the submission of video recordings from clinical practice."
	Addition of a written examination	#1: "I think it would be good to have a written component to ensure knowledge standards. For example, like US board certification."
	Addition of a practical examination	#1: "I think it would be good to include more practical assessment, OSCE-style."
Alternative perspectives on examination format and training	Desire for an examination format that reflects processes and competencies	#7: "I would really like the examination format itself to be reconsidered. Moving away from this conversational format towards something else that reflects processes and competencies."
	Desire for longitudinal assessment	#2: "I would like us to have a conversation with colleagues every six months, to practise skills and review the skills they have learned. That we really assess certain knowledge or tasks regularly over the course of these five years. It is completely clear to me that this is a huge effort and a process that cannot be implemented overnight, but in my view it could significantly contribute to improving both quality and satisfaction during training."
	Focus on improving the quality of training rather than focusing on the exam	#6: "And I would focus more on improving the quality of the training afterwards, rather than making the exam punitive."

"What we can say is that those who participated in a competence centre generally perform very well in the examinations. This does not mean that someone who did not attend a competence centre cannot be excellent, but those who performed poorly were usually not affiliated with a competence centre." (#11)

Discussion

This qualitative study explored examiners' perspectives on the current specialist examination in general practice, including its perceived strengths and weaknesses, potential areas for further development and the influence of competence centres for postgraduate training. In addition, the study investigated examiners' motivations for participating in assessment activities and their views on the qualifications required for examiners.

The demographic changes within the medical workforce [22] are also reflected among specialist examiners. Given the increasing number of specialist certifications in general practice, a growing need for the recruitment and qualification of new examiners can be anticipated. Participants primarily reported intrinsic and discipline-related motivations for serving as examiners. Adequate financial

compensation for lost income may facilitate the recruitment of additional examiners.

The findings reveal an ambivalent attitude towards the current specialist examination. Although participants agreed that the examination enables them to determine whether candidates meet the minimum standard required for specialist certification, they simultaneously acknowledged substantial limitations regarding objectivity, reliability, validity and contextual sensitivity. These findings are consistent with international research on oral examinations. In their review, Memon et al. [23] demonstrated that oral examinations remain deeply embedded within postgraduate medical training in many countries, although some countries have replaced them with alternative formats because of their limited psychometric properties. In Switzerland, for example, specialist certification in general internal medicine includes a written examination consisting of 120 multiple-choice questions, which increases assessment objectivity [24]. Burch et al. [25] questioned the extent to which specialist certification examinations can serve as internationally comparable indicators of physician competence. According to Miller's pyramid [15], oral examinations predominantly assess cognitive knowledge and clinical reasoning ("knows" and "knows how"), whereas higher levels of competence relating to actual clinical performance ("shows how" and "does") cannot be adequately captured in an oral format alone. However, the assessment of independent professional practice represents the central purpose of specialist certification as the culmination of postgraduate training. Participants expressed highly heterogeneous views regarding possible examination reforms. Overall, there appeared to be limited motivation among examiners to fundamentally change the current assessment format. Similarly, Fink et al. [17] reported little interest among trainees in general practice in modifying the existing examination. At present, there appears to be no substantial bottom-up initiative for reform, either among examiners or among physicians in postgraduate training. Nevertheless, participants identified several short-term and easy to implement measures for improving the current examination format, including examiner training, regular exchange among examiners (e.g. examiner quality circles) and the development of jointly agreed assessment frameworks and expected standards. Such measures may improve administration objectivity, scoring objectivity and interpretative objectivity. Beckers et al. [26] demonstrated considerable variation in examination content and competency levels in specialist examinations in internal medicine and recommended the implementation of examination blueprints similar to those used in the United States to ensure comparable examination conditions for all candidates. In several European countries, examiners are already required to participate in formal examiner training before conducting assessments [27]. Examiner training programmes implemented in Southern Baden have shown high levels of acceptance among participants [28]. Examiners considered the definition of minimum standards and harmonisation of expected performance levels and assessment criteria essential for improving examination validity and reliability.

Based on the qualification criteria identified by the examiners in the study, it is possible to derive a profile of graduates of examiner training programs, which can serve as the basis for further developing corresponding teaching and learning objectives for such training programs. Examiner training should address both practical competencies in conducting assessments and knowledge of different assessment formats, while simultaneously providing opportunities for critical reflection on the balance between practical feasibility and educational assessment standards. The active involvement of examiners will be essential for any sustainable restructuring of the examination system.

Analogous to the requirement for newly appointed postgraduate trainers in Germany to attend train-the-trainer courses [29], mandatory compact examiner training programmes for newly appointed examiners could be considered. Such programmes are already established in countries including Denmark, the Netherlands and the United Kingdom [27]. Targeted training may also reduce uncertainty and administrative burden in cases of examination failure, while predefined assessment frameworks may provide a transparent and defensible basis for pass-fail decisions. The 2018 German model specialty training regulations (MWBO) describe primarily formal aspects of specialist examinations [1]. Given the transition toward a competency-based approach in the new model speciality training regulations, it would be beneficial to align the examination format accordingly with this competency-based approach. Proposals for structured oral examinations aimed at improving quality and standardisation have already been published [30]. From an educational perspective, formative assessments ("assessment for learning") are better suited to supporting continuous learning processes than summative examinations ("assessment of learning"). In psychotherapy training in Germany, for example, there is an intermediate examination and a final examination. The final examination consists of a written multiple-choice test and an oral case study examination [https://www.gesetze-im-internet.de/psychthappro/BJNR044800020.html]. Internationally, formative assessment approaches have become established in postgraduate training in general practice in countries such as Denmark and the Netherlands [31], [32], where summative final examinations are largely absent. In France, trainees document their competency development in portfolios that are supervised by educational mentors (*référents pédagogiques*) [33]. Similarly, Khan et al. [34] demonstrated advantages of formative assessment for enhancing individual learning trajectories in postgraduate urology training. Competencies should therefore be continuously assessed and supported through formative workplace-based assessment methods, such as Mini Clinical Evaluation Exercises (Mini-CEX), Direct Observation of Procedural Skills (DOPS) or portfolio assessments [35]. In recent years, Entrustable Professional Activities (EPAs) have emerged as promising instru-

ments for capturing higher levels of clinical performance within Miller's framework [36]. However, evidence from the Netherlands suggests that EPAs may currently be more valuable as learning tools than as high-stakes assessment instruments in general practice training [37]. Thiessen et al. [38] proposed a best-practice model for specialist examinations based on the triangulation of multiple assessment methods, blueprinting, predefined questions and expected standards, quality assurance procedures and educational impact on candidates. Such approaches inevitably require greater financial and personnel resources [27].

Overall, the findings reveal a tension between awareness of the shortcomings of the current examination format and motivation for fundamental reform. From an assessment-theoretical perspective, further development appears advisable in order to achieve better alignment between competency goals and assessment methods. Given the importance of assessment both for learning processes and for quality assurance in healthcare, such adaptation appears educationally necessary. At the same time, participants repeatedly emphasised practical implementation challenges, particularly the limited availability of examiners and concerns regarding feasibility.

A longitudinal formative assessment programme based on portfolios and logbooks, including practical assessment components, could also strengthen the quality and efficiency of postgraduate training in Germany. Although elements of this are currently addressed through educational supervisors and logbooks, an independent assessment body responsible for triangulating competence assessments may provide additional quality assurance. As many candidates enter general practice training through career transitions or lateral entry [39], formative assessments could provide early feedback regarding their professional reorientation and guide subsequent learning goals. International examples further demonstrate the value of progress testing in postgraduate education [40], enabling trainees to monitor learning progress and identify future educational needs. Participation in formative assessments for feedback purposes could therefore also be considered on a voluntary basis.

A proportion of participants attributed better examination performance and lower failure rates to trainees who had participated in CCPT programmes. These observations may be interpreted as potential effects of the Competence Centres for Postgraduate Training. However, selection effects involving particularly motivated trainees, as well as other confounding factors such as stronger professional networks, must also be considered. Particularly noteworthy was the frequently reported observation that candidates who failed the examination had often not participated in CCPT programmes.

Strengths and limitations

This study included examiners with both extensive and limited examination experience, thereby strengthening the credibility and richness of the findings. Nevertheless, several limitations should be acknowledged. As a qualitative study, the findings do not claim completeness or representativeness. Examiners from all German federal states were not included and the total number of general practice specialist examiners in Germany is currently unknown. However, the sample included both less experienced and highly experienced examiners and is likely to reflect the composition of examination committees reasonably well. A potential selection bias must be considered, as recruitment was partly conducted through existing professional networks and predominantly attracted already well-connected examiners. It is conceivable that examiners with stronger academic affiliations may be more receptive to innovative assessment formats. In addition, the possibility of social desirability bias within the interviews cannot be excluded.

The reported effects of the competence centres for postgraduate training should be interpreted as exploratory, as not all participants were aware of candidates' participation in these programmes and therefore we may not have been able to explore all relevant perspectives. Future mixed-methods studies should further investigate both the feasibility and educational value of potential reforms to the specialist examination as well as the effects of competence centres for postgraduate training.

Conclusions

Examiners demonstrated considerable awareness of the limitations of the current specialist examination and critically reflected on its objectivity, reliability and validity in relation to the complexity of general practice. Although views regarding desirable reforms varied, participants considered the current examination capable of assessing the minimum standard required for specialist certification. From an educational perspective, clearer specification of the objectives of the specialist examination, improved alignment between the competency goals of the model specialty training regulations and assessment formats and the introduction of longitudinal assessment approaches incorporating practical assessment components should be considered. As immediately feasible measures, nationwide examiner training, regular exchange among examiners and the implementation of shared assessment frameworks appear particularly promising.

Notes

Ethics approval and consent to participate

The study adhered to the Declaration of Helsinki. All participants gave written consent to participate in the study and to the pseudonymised use of their data. The studies were approved by University of Heidelberg, Medical University, Ethic Committee Alte Glockengießerei 11/169115 (S711/2023). The studies were conducted in accordance

with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Availability of data and materials

The datasets presented in this article are not readily available because as required by the European Data Protection law the participants were informed of the privacy policy and agreed that the pseudonymized data would be available to the project team and would be stored at the study center. Request to access the datasets should be directed to Simon Schwill (simon.schwill@med.uni-heidelberg.de).

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. The KWBW Verbundweiterbildungplus® is supported by public funding under Section 75a of the German Social Code V, Annex

Author contributions

ACD was integrated in idea, in conceptualisation and methodology, to acquisition, analysis and interpretation of data and writing the original draft. CB was integrated into acquisition, analysis and interpretation of data. MR was integrated in methodology, to analysis of data. AA was integrated to interpretation of data and reviewing the manuscript. SS was integrated in idea and funding for the study, in conceptualisation and methodology, to analysis and interpretation of data and writing the original draft.

Authors' ORCIDs

- Andreas Christian Dreher: [0009-0003-6688-8807]
- Christine Becker: [0009-0009-4802-5761]
- Marco Roos: [0000-0003-1596-5908]
- Attila Altiner: [https://orcid.org/0000-0002-2429-933X]
- Simon Schwill: [0000-0002-0954-2194]

Acknowledgements

We like to thank all participating GP examiners, all KWBW practices and partner for their participation. The study was presented at the annual conference of the DACH society for medical education in the health professions, Düsseldorf 2025.

Competing interests

ACD, CB and SS declare to be integrated in the organisation of the postgraduate training programme KWBW. All authors declare that they have no competing interests.

Attachments

Available from https://doi.org/10.3205/zma001872

1. Attachment_1.pdf (417 KB)
   COREQ (COnsolidated criteria for REporting Qualitative research) checklist
2. Attachment_2.pdf (268 KB)
   Specialist examination/effects of the competency-based curriculum (KW) – specialist examiners
3. Attachment_3.pdf (168 KB)
   Illustrative quotes

References

1. Bundesärztekammer. Postgraduate training curriculum: Muster-Weiterbildungsordnung 2018 in der Fassung vom 26.06.2021. Berlin: Bundesärztekammer; 2021.
2. Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. Med Teach. 2007;29(7):642-647. DOI: 10.1080/01421590701746983
3. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrot V, Roberts T. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):206-214. DOI: 10.3109/0142159X.2011.551559
4. Bundesärztekammer. Ärztestatistik zum 31. Dezember 2024. Berlin: Bundesärztekammer; 2024. Zugänglich unter/available from: https://www.bundesaerztekammer.de/fileadmin/user_upload/BAEK/Ueber_uns/Statisik/AErztestatistik_2024.pdf
5. Landesärztekammer Baden-Württemberg. Neue Ärztestatistik für Baden-Württemberg. Stuttgart: Ärztekammer Baden-Württemberg; 2024. Zugänglich unter/available from: https://www.aerztekammer-bw.de/neue-aerztestatistik-fuer-baden-wuerttemberg-f14fcf0d7b2267dc
6. Dißmann P. Europäische Facharztprüfung für Notfallmedizin. Notfall Rettungsmed. 2014;17:697-700. DOI: 10.1007/s10049-014-1937-5
7. Schwill S, Magez J, Joos S, Steinhäuser J, Ledig T, Rubik A, Niebling W, Szecsenyi J, Flum E. New paths in post-graduate medical training in general practice - 8 years of experience with the pilot project Verbundweiterbildungplus Baden-Württemberg. GMS J Med Educ. 2017;34(5):Doc62. DOI: 10.3205/zma001139
8. Elmer A, Grifka J. Vergleich von Prüfungsmethoden in der medizinischen Ausbildung. Gesundheitswesen (Suppl Med. Ausbild). 1998;15(Suppl1):14-17. Zugänglich unter/available from: https://gesellschaft-medizinische-ausbildung.org/files/ZMA-Archiv/1998/1/Elmer_A,_Grifka_J.pdf
9. Artino AR Jr, Durning SJ, Creel AH. Reliability and validity in educational measurement. Acad Med. 2010;85(9):1545. DOI: 10.1097/ACM.0b013e3181edface
10. Wormald BW, Schoeman S, Somasunderam A, Penn M. Assessment drives learning: an unavoidable truth? Anat Sci Educ. 2009;2(5):199-204. DOI: 10.1002/ase.102
11. van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. Med Educ. 2005;39(3):309-317. DOI: 10.1111/j.13652929.2005.02094.x

12. Yazbeck Karam V, Park YS, Tekian A, Youssef N. Evaluating the validity evidence of an OSCE: results from a new medical school. BMC Med Educ. 2018;18(1):313. DOI: 10.1186/s12909-018-1421-x

13. Pell G, Fuller R, Homer M, Roberts T; International Association for Medical Education. How to measure the quality of the OSCE: a review of metrics – AMEE Guide No. 49. Med Teach. 2010;32(10):802-811. DOI: 10.3109/0142159X.2010.507716

14. ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? Acad Med. 2007;82(6):542-547. DOI: 10.1097/ACM.0b013e31805559C7

15. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9):S63-S67. DOI: 10.1097/00001888-199009000-00045

16. Biggs J. Enhancing teaching through constructive alignment. High Educ. 1996;32(3):347-364. DOI: 10.1007/BF00138871

17. Fink M, Roos M, Schwill S, Steinhäuser J. Facharztprüfung Allgemeinmedizin: Quo vadis? ZFA (Stuttgart). 2025;101:316-321. DOI: 10.1007/s44266-025-00377-y

18. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349-357. DOI: 10.1093/intqhc/mzm042

19. Kern DE, Thomas PA, Hughes MT. Curriculum development for medical education: a six-step approach. 2nd ed. Baltimore: Johns Hopkins University Press; 2009.

20. Van Someren M. The think aloud method. London: Academic Press; 1994.

21. Kuckartz U, Rädiker S. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. 5. überarbeitete Auflage. Weinheim: Beltz Juventa; 2022.

22. Bundesärztekammer. Arztzahlstatistik: Jede/r vierte berufstätige Ärztin/Arzt ist 60 Jahre und älter. Berlin: Bundesärztekammer; 2023. Zugänglich unter/available from: https://www.bundesaerztekammer.de/presse/aktuelles/detail/arztzahlstatistik-zeigt-jede-r-vierte-berufstaetige-aerztin-arzt-ist-60-jahre-und-aelter

23. Memon MA, Joughin GR, Memon B. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. Adv Health Sci Educ Theory Pract. 2010;15(2):277-289. DOI: 10.1007/s10459-008-9111-9

24. Schweizerisches Institut für ärztliche Weiter- und Fortbildung (SIWF). Fachärztin oder Facharzt für Allgemeine Innere Medizin. Weiterbildungsprogramm. Bern: Schweizerisches Institut für ärztliche Weiter- und Fortbildung; 2022. Zugänglich unter/available from: https://www.siwf.ch/files/pdf21/aim_version_internet_d.pdf

25. Burch VC, Norman GR, Schmidt HG, van der Vleuten CP. Are specialist certification examinations a reliable measure of physician competence? Adv Health Sci Educ Theory Pract. 2008;13(4):521-533. DOI: 10.1007/s10459-007-9063-5

26. Beckers MA, Braun LT, Epstein N, Fischer MR, Schmidmaier R. Die Facharztprüfung Innere Medizin: Eine qualitative und quantitative Analyse von Gedächtnisprotokollen. Dtsch Med Wochenschr. 2020;145(2):e10-e17. DOI: 10.1055/a-0970-6348

27. Flum E, Maagaard R, Godycki-Cwirko M, Scarborogh N, Scherpbier N, Ledig T, Roos M, Steinhäuser J. Assessing family medicine trainees – what can we learn from the European neighbours? GMS Z Med Ausbild. 2015;32(2):Doc21. DOI: 10.3205/zma000963

28. Streitlein-Böhme I, Niebling W, Fabry G, Böhme K. Three years of experience with a workshop for medical specialty examiners in South Baden: A project report and initial evaluation results. GMS J Med Educ. 2017;34(5):Doc55. DOI: 10.3205/zma001132

29. Bundesärztekammer. Beschlussprotokoll des 128. Deutschen Ärztetags. Berlin: Bundesärztekammer; 2024. Zugänglich unter/available from: https://www.bundesaerztekammer.de/fileadmin/user_upload/BAEK/Aerztetag/128.DAET/2024-05-10_Beschlussprotokoll_neu.pdf

30. Glaab T, Böhme K, Griesinger J, Forster A, Streitlein-Böhme I. Die mündliche Facharztprüfung: Ist das aktuelle Format noch zeitgemäß? Dtsch Med Wochenschr. 2021;146(18):1226-1229. DOI: 10.1055/a-1506-1280

31. IGES Institut. Untersuchung zur Förderung der Weiterbildung Allgemeinmedizin im internationalen Vergleich. Berlin: IGES; 2023. Zugänglich unter/available from: https://www.iges.com/ergebnisse/projekte/2023/weiterbildung-allgemeinmedizin/index_ger.html

32. Miani C, Hinrichs S, Pitchforth E, Bienkowska-Gibbs T, Disbeschl S, Roland M, Nolte E. Best Practice: Medizinische Aus- und Weiterbildung aus internationaler Perspektive. Cambridge (UK): Rand Europe; 2015. Zugänglich unter/available from: https://www.rand.org/content/dam/rand/pubs/research_reports/RR600/RR622z1/RAND_RR622z1.pdf

33. Held LA. Allgemeinmedizin in Studium und Weiterbildung in Frankreich. ZFA (Stuttgart). 2021;2:78-83. DOI: 10.3238/zfa.2021.0078-0083

34. Khan RN, Siddiqui NA. The Use of Formative Assessment in Postgraduate Urology Training: A Systematic Review. Cureus. 2022;14(7):e27162. DOI: 10.7759/cureus.27162

35. Driessen E, Scheele F. What is wrong with assessment in postgraduate training? Lessons from clinical practice and educational research. Med Teach. 2013;35(7):569-574. DOI: 10.3109/0142159X.2013.798403

36. O'Dowd E, Lydon S, O'Connor P, Madden C, Byrne D. A systematic review of 7 years of research on entrustable professional activities in graduate medical education, 2011-2018. Med Educ. 2019;53(3):234-249. DOI: 10.1111/medu.13792

37. Bonnie LHA, Visser MRM, Bont J, Kramer AWM, van Dijk N. Trainers' and trainees' expectations of entrustable professional activities (EPAs) in a primary care training programme. Educ Prim Care. 2019;30(1):13-21. DOI: 10.1080/14739879.2018.1532773

38. Thiessen N, Fischer MR, Huwendiek S. Assessment methods in medical specialist assessments in the DACH region - overview, critical examination and recommendations for further development. GMS J Med Educ. 2019;36(6):Doc78. DOI: 10.3205/zma001286

39. Dreher AC, Ko J, Becker C, Förster C, Jähnig T, Bischoff M, Stengel S, Schwill S. Analyse der Zielgruppen in der Weiterbildung Allgemeinmedizin: Ergebnisse einer Kohortenstudie im Kompetenzzentrum Weiterbildung Baden-Württemberg (KWBW) [Analyzing target audience in family medicine specialty training: Results of a cohort study at the Competence Center for Postgraduate Medical Training]. Z Evid Fortbild Qual Gesundhwes. 2025:197;55-61. DOI: 10.1016/j.zefq.2025.06.004

40. Dijksterhuis MG, Scheele F, Schuwirth LW, Essed GG, Nijhuis JG, Braat DD. Progress testing in postgraduate medical education. Med Teach. 2009;31(10):e464-e468. DOI: 10.3109/01421590902849545

Corresponding author:
PD Dr. med. Simon Schwill, MME
University Hospital Heidelberg, Department for Primary Care and Health Services Research, Im Neuenheimer Feld 130.3, D-69120 Heidelberg, Germany, Phone: +49 (0)6221/56-39379
simon.schwill@med.uni-heidelberg.de

Please cite as
Dreher AC, Becker C, Roos M, Altiner A, Schwill S. The general practice specialist examination – examiners' perspectives on the status quo and future development: An exploratory qualitative study. GMS J Med Educ. 2026;43(6):Doc78.
DOI: 10.3205/zma001872, URN: urn:nbn:de:0183-zma0018723

This article is freely available from
https://doi.org/10.3205/zma001872

Received: *2025-10-29*
Revised: *2026-03-30*
Accepted: *2026-04-30*
Published: *2026-09-15*

Copyright
©2026 Dreher 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/.

Die Facharztprüfung Allgemeinmedizin – Perspektiven von Prüfenden auf Status Quo und Weiterentwicklung: Eine explorative, qualitative Studie

Andreas Christian Dreher[1]
Christine Becker[1]
Marco Roos[2]
Attila Altiner[1]
Simon Schwill[1]

1 Universitätsklinikum Heidelberg, Abteilung für Allgemeinmedizin und Versorgungsforschung, Heidelberg, Deutschland

2 Universität Augsburg, Medizinische Fakultät, Abteilung für Allgemeinmedizin, Augsburg, Deutschland

Zusammenfassung

Hintergrund: Die Facharztprüfung beeinflusst in hohem Maße die Versorgungsqualität. Ziel der Studie war die Perspektiven von Prüfenden zu Status quo, Motivation, Vor- und Nachteilen, Veränderungsvorschlägen sowie Effekten der Kompetenzzentren Weiterbildung (KW) zu explorieren.
Material und Methoden: Prüfende wurden deutschlandweit rekrutiert und anhand eines selbstentwickelten Fragebogens in semi-strukturierten Video-Interviews befragt. Die Aussagen wurden Wort für Wort transkribiert und inhaltanalytisch ausgewertet.
Ergebnisse: Es nahmen 12 Prüfende aus 6 Bundesländern teil. Die durchschnittliche Interviewdauer betrug 33 Minuten (Min. 24 min, Max. 57 min). Als Vorteile wurde das Einzelprüfungssetting, die Sicherung des Mindeststandards sowie die Zeit- und Kosteneffizienz hervorgehoben. Als nachteilig wurde die begrenzte Aussagekraft mündlicher Prüfungen für die komplexen Handlungskompetenzen an Patient:innen gesehen. Der wahrgenommene Veränderungsbedarf variierte zwischen den Prüfenden. Vorgeschlagen wurden Schulungen, Erwartungshorizonte und praktische Prüfungsanteile. Zudem wurden longitudinale Assessments benannt.
Diskussion: Prüfende sehen Schwächen der Prüfung bezüglich Objektivität, Reliabilität, Validität und Kontextsensibilität in Bezug auf die Komplexität allgemeinmedizinischer Arbeitsweise, halten aber die Einschätzung des Mindeststandards der Facharztqualifikation für möglich. Es zeigt sich ein Spannungsfeld zwischen praktischer Umsetzbarkeit und prüfungsdidaktischem Anspruch. Potenziale zur Weiterentwicklung bestehen in Prüfendenschulungen, regelmäßigem Austausch, gemeinsamen Erwartungshorizonten, einer geschärften Zieldefinition und longitudinalem Assessment (Portfolio mit regelmäßigen Teilprüfungen).
Schlussfolgerung: Die Facharztprüfung weist Entwicklungspotenzial auf. Vergleichbare Erwartungshorizonte, wie sie bei Prüfendenschulungen erlernt werden, strukturieren die Prüfung und könnten einfach implementiert werden.

Schlüsselwörter: Facharztprüfung, Prüfendenschulung, Weiterbildung, Kompetenzzentren Weiterbildung

Einleitung

Die Facharztprüfung Allgemeinmedizin steht am Ende von fünf Jahren Weiterbildung in Vollzeitäquivalent [1]. In Deutschland sind die Landesärztekammern für die Durchführung der Prüfungen zuständig, sie wird von den Ärztekammern organisiert und dient der formalen Anerkennung der fachärztlichen Kompetenz. Die Facharztprüfung ist die letzte Prüfung in der ärztlichen Laufbahn und das Bestehen erlaubt die eigenständige und eigenverantwortliche fachärztliche Tätigkeit. Sie ist ein Instrument zur Sicherstellung der fachärztlichen Qualifikation mit Auswirkung auf die medizinische Versorgungsqualität und Patientensicherheit [2], [3]. Es handelt sich in der Regel um eine mündliche Prüfung über mindestens 30 Minuten mit zwei (Fach-)Prüfenden und einem Prüfungsvorsitz. Klassischerweise handelt es sich um ein unstrukturiertes „kollegiales Fachgespräch", das sich in Deutschland seit Jahrzehnten etabliert hat. Die Umsetzung kann je nach Region, Fachgebiet und individueller Prüfungspraxis variieren.

Die Einführung der fünfjährigen Weiterbildung und der Facharztprüfung für Allgemeinmedizin markiert einen zentralen Schritt zur fachlichen Gleichstellung der Allgemeinmedizin mit anderen medizinischen Fächern. In den letzten Jahren ist die Anzahl der Facharztprüfungen Allgemeinmedizin gestiegen [4] und ist beispielsweise aktuell in Baden-Württemberg die häufigste abgenommene Facharztprüfung [5]. Die Nichtbestehen-Quoten in der Facharztprüfung sind in Deutschland im einstelligen Prozentbereich, während beispielsweise die europäische Facharztprüfung Notfallmedizin Durchfallquoten bis 64% hat [6]. Ergänzend zur praktischen Weiterbildung unter Supervision existieren in Deutschland Kompetenzzentren Weiterbildung (KW) zur Steigerung der Qualität und Effizienz der Weiterbildung Allgemeinmedizin [7]. Hier wird die Weiterbildung durch Seminare mit Curriculum, Mentoring sowie der Teilnahme an einem Weiterbildungsverbund unterstützt.

Generell sollten Prüfungen objektiv, valide, reliabel [8], [9] und kontextsensibel sein [3]. In der medizinischen Ausbildung wird der Prüfung zur Steuerung des Lernprozesses eine große Bedeutung zugeschrieben (*assessment drives learning*) [10], [11]. Zudem haben kompetenzorientierte, praktische und standardisierte Prüfungsformate an Bedeutung gewonnen [12], [13] und konnten sich an vielen Universitäten etablieren. Konzepte von *Competency-Based Medical Education (CBME)* und *Entrustable Professional Activities (EPAs)* rücken stärker in den Fokus [14]. Auch die Musterweiterbildungsordnung (MWBO) hat sich in der Neuauflage [1] kompetenzorientiert ausgerichtet, aber das Prüfungsformat wurde in seiner ursprünglichen Form beibehalten. Die aktuelle Facharztprüfung ist aus testtheoretischer Sicht nicht ausreichend, um Kompetenzorientierung und praktische Fertigkeiten valide abzubilden [15]. Im Sinne des Constructive Alignments [16] sollten Lern- und Kompetenzziele, Weiterbildungsaktivitäten und Prüfungsformate aufeinander abgestimmt sein, um zielgerichtetes Lernen zu fördern. Gegenwärtig ist von einem begrenzten Alignment von MWBO und der derzeit meist unstrukturierten mündlichen Prüfung auszugehen. Aus prüfungsdidaktischer Sicht ist eine grundlegende Weiterentwicklung der Facharztprüfung für ein besseres Alignment mit den Kompetenzzielen der neuen MWBO angezeigt.

Die Facharztprüfung bildet den Abschluss der allgemeinmedizinischen Weiterbildung und beeinflusst damit die Versorgungsqualität. Eine Umfrage unter Ärzt*innen in Weiterbildung und Fachärzt*innen für Allgemeinmedizin ergab keinen Wunsch nach Veränderung [17]. Ziel der explorativen Studie war es zu erfassen, wie Prüfende den Status Quo der aktuellen Prüfung bewerten, welche Motivationen ihrer Prüfungstätigkeit zugrunde liegen, welche Vor- und Nachteile sie dem bestehenden Format zuschreiben, welche Weiterentwicklungsbedarfe sie identifizieren, welche Qualifikationen für Prüfende notwendig sind und inwiefern Effekte der Kompetenzzentren Weiterbildung (KW) auf die Prüfung wahrgenommen werden.

Material und Methoden

Art der Studie und Setting

Es wurde eine qualitative Studie mit semi-strukturierten Interviews mit Prüfenden für Allgemeinmedizin durchgeführt. Der qualitative Ansatz wurde mit dem Ziel der tiefergehenden Erfassung von Erfahrungen und wahrgenommener Herausforderungen in ihrer Komplexität und Kontextgebundenheit gewählt. Innerhalb eines Zeitraums von 16 Wochen wurden von Dezember 2024 bis April 2025 die Rekrutierung, die Interviews und die anschließende Datenanalyse durchgeführt. Die Berichterstattung basiert auf den konsolidierten Kriterien für die Berichterstattung über qualitative Forschung (COREQ) (siehe Anhang 1) [18].

Teilnehmende und Rekrutierung

Alle Befragten waren Prüfende für Allgemeinmedizin und wurden via E-Mail kontaktiert. Es bestand keine öffentlich zugängliche Liste von Prüfenden in Deutschland oder Netzwerk von Prüfenden, sodass die Rekrutierung nach *purposive sampling* gezielt bezüglich Bundesland, Prüfungserfahrung und Geschlecht erfolgte.

Prüfende wurden sowohl über bestehende berufliche Netzwerke (Kompetenzzentren Weiterbildung, Fachkongresse) als auch durch gezielte Identifikation bislang nicht bekannter Personen anhand von Prüfungsprotokollen kontaktiert. Einzelne Teilnehmende waren den Forschenden aus dem beruflichen Umfeld bekannt. Es bestand jedoch kein hierarchisches oder anderweitiges Abhängigkeitsverhältnis zwischen Forschenden und Teilnehmenden. Es wurden 40 Prüfende kontaktiert, wovon sich 13 zurückmeldeten und 12 bei einem Interview teilnahmen.

Entwicklung des Instruments, Interviewleitfaden

Nach einer Literaturrecherche (NIH/PubMed®: „specialist exam“, „board certification“, „resident assessment“, „postgradual medical education assessment“) wurde ein Interviewleitfaden von einer Gruppe von vier Forschenden (ACD, CB, MR, SS) entwickelt, die sowohl in qualitativer Forschung als auch in Allgemeinmedizin und Didaktik erfahren waren.
Der Kern-Zyklus [19] diente als theoretisches Framework zur Strukturierung der Studie, da er die systematische Verknüpfung von Bedarfsanalyse, Zieldefinition, Lehr-/Lernaktivitäten und Assessment beschreibt. Sowohl ein Needs Assessment der Zielgruppe der Prüfenden als auch eine Einordnung entlang des Kern-Zyklus im Kontext der Weiterbildung wurden in der Entwicklung des Interviewleitfadens berücksichtigt.
Die Interviews waren leitfadengestützt, um eine systematische Abdeckung der zentralen Themenfelder zu gewährleisten. Gleichzeitig wurden freie Assoziationen und vertiefende Rückfragen im Interview ermöglicht.
Der Interviewleitfaden wurde mittels Think-Aloud-Technik validiert [20]. In dieser Methode verbalisieren die Teilnehmenden ihre Gedankenprozesse, kognitiven Strategien und Entscheidungsvorgänge. Anschließend wurde der Interviewleitfaden mit zwei Interviews pilotiert und geringfügig angepasst (siehe Anhang 2). Der Interviewleitfaden umfasst Fragen zur Soziodemographie und Erfahrung als Prüfende, Einschätzungen der aktuellen Facharztprüfung, Veränderungsperspektiven sowie Fragen zu möglichen Effekten durch die Kompetenzzentren Weiterbildung.

Datenerhebung

Die Interviews wurden über Zoom® (San Jose, California USA, Version 6.3.0) von ACD (Facharzt für Psychosomatische Medizin und Psychotherapie, Arzt in fortgeschrittener Weiterbildung Allgemeinmedizin, Prüfer für Psychologische Psychotherapeut*innen und Kinder- und Jugendpsychotherapeut*innen, mehrjährige Erfahrung in Interviewführung und qualitativer Datenanalyse) durchgeführt. Eingangs wurden die Teilnehmenden über den Hintergrund der Studie und die Rolle des Forschenden informiert. Keine dritte Person (Nicht-Teilnehmende) war am Arbeitsplatz des Forschers und im privaten Umfeld der Prüfenden anwesend. Es wurde eine Audioaufnahme aufgenommen. Feldnotizen wurden angefertigt, um relevante Beobachtungen während der Interviewdurchführung zu dokumentieren. Die Transkripte wurden den Prüfenden nicht zur Korrektur gegeben.

Qualitative Datenanalyse

Die Interviews wurden Wort für Wort transkribiert und papierbasiert analysiert. Die Transkripte wurden gesichtet und induktiv-deduktiv angelehnt an Kuckartz kategorisiert [21] und von zwei Forschern (ACD, SS) unabhängig voneinander analysiert. Anschließend wurden Kategorien und Zitate verglichen und bei Bedarf angepasst. Bei Unklarheiten wurde zu dritt (CB) das Material gesichtet, um einen Konsens zu finden. Während der Auswertung fand kontinuierlich eine Reflexion von persönlichen, interpersonalen, methodischen und kontextbezogenen Faktoren statt, welche die Analyse beeinflussen könnten.

Ergebnisse

Interviewdurchführung

Es wurden n=12 Interviews zwischen Dezember 2024 und April 2025 durchgeführt. Die durchschnittliche Interviewdauer betrug 33 Minuten (Minimum 24 min, Maximum 57 min). Die Durchführung erfolgte ohne technische Probleme. Kein Interview musste wiederholt werden. Das Kategoriensystem wurde nach zwei Interviews überprüft und keine weiteren Anpassungen waren notwendig. In den letzten Interviews ergaben sich keine neuen Aspekte (Codes), sodass von einer *meaning saturation* ausgegangen wurde und keine weitere Rekrutierung erfolgte.

Soziodemographische Daten und Prüfungserfahrung

Die soziodemographischen Daten sind in Tabelle 1 dargestellt. Alle an der Studie beteiligten Prüfenden waren Fachärzt*innen (FÄ) für Allgemeinmedizin und besaßen relevante Expertise in der Durchführung von Facharztprüfungen. In den befragten Bundesländern gab es den Unterschied, ob der Prüfungsvorsitz fachfremd oder facheigen ist. Zudem war es zwischen den Bundesländern nicht einheitlich, ob die Prüfung auf Tonband aufgenommen wurde.

Motivation für Prüfendentätigkeit

Die Motivation für die Tätigkeit als Prüfende ist überwiegend von idealistischen und intrinsischen Motiven getragen und in Tabelle 2 dargestellt. Die Aufwandsentschädigung, welche den Praxisausfall nicht kostenneutral kompensiert, wurde als mögliche Barriere in der Rekrutierung neuer Prüfender gesehen.

Vor- und Nachteile aktueller Prüfungsform

Die berichteten Vor- und Nachteile der aktuellen Prüfungsform sind in Tabelle 3 gegenübergestellt. In Anhang 3 sind repräsentative Ankerzitate dargestellt. Die aktuelle Facharztprüfungsform hat aus Sicht der meisten Prüfenden organisatorische Stärken und inhaltliche Vorteile. Als großen Vorteil sieht die Mehrheit der Prüfenden die gute Durchführbarkeit der Prüfung bei geringem Zeitaufwand und praktischer Umsetzbarkeit. Die Prüfung durch Fachkolleg*innen ermögliche eine fachlich fundierte Qualitäts-

Tabelle 1: Soziodemographische Daten der Facharztprüfenden, die an der qualitativen Studie teilnahmen (N=12)

Alter	**Jahre (Median, IQR)**	**62 (56, 66)**
Geschlecht	weiblich	7
	männlich	5
Bundesland	Baden-Württemberg Bayern Bremen Hessen Niedersachsen Sachsen	3 2 1 2 1 3
Facharztbezeichnung	Fachärzt*in für Allgemeinmedizin	12
Jahre Tätigkeit als Prüfende*r	Jahre (SD) Jahre (Median, IQR)	13,42 (9,91) 12 (5,20)
Anzahl abgenommener Prüfungen	Anzahl Prüfungen (SD) Anzahl Prüfungen (Median, IQR)	200,67 (82,5) 178 (136,254)

SD=Standardabweichung, IQR=Interquartile Range

Tabelle 2: Motivationsgründe für die Tätigkeit als Facharztprüfende

Motivation aus Gründen für das Fach	Qualitätssicherung	*#1: „Ich halte es für eine wichtige Tätigkeit im Rahmen der Qualitätssicherung Allgemeinmedizin."*
	Beitrag für die Weiterbildung Allgemeinmedizin leisten	*#2: „Also mich motiviert zum einen mein Wunsch, einen Beitrag für die Weiterbildung zu leisten und dazu gehört für mich eben auch diese in Anführungszeichen Last auf sich zu nehmen, Prüfungen abzuhalten, weil es für mich die Weiterbildung auch rund macht."*
Intrinsische/persönliche Motivation	Ehre/Überzeugung	*#7: „für mich ist es persönlich eine Ehre, Facharztprüfungen abzunehmen (…) Dass man da am Ende quasi der ganzen Strecke von Ausbildung zur Weiterbildung steht und sieht, was rauskommt."* *#8: „Also aus Überzeugung, dass jemand das machen muss"*
	Notwendigkeit für fachlich Up-to-Date-Bleiben	*#9: Hauptsächlich war es aber tatsächlich der Grund, dass ich dann gezwungen bin, mich auch wieder medizinisch auf den neuesten Stand zu bringen."*
Gesellschaftlich-strukturelle Motive	Besetzung der Prüfungskommission unter Fachgebietsaspekten (berufspolitisch)	*#10: „und dass Allgemeinärzte Allgemeinärzte prüfen"*
	Besetzung der Prüfungskommission unter Gender-Aspekten	*#10: „das Gender-Thema, die Ärztinnen werden weiblicher und dann braucht es auch weibliche Prüferinnen"*

kontrolle auch bezüglich des Prüfens von Haltungen. Das Setting einer persönlichen Einzelprüfung und den Gestaltungsspielraum als Prüfende in der Fragen- und Themenauswahl wurde von einigen Prüfenden als Vorteil gesehen. Gleichzeitig wurden verschiedene Nachteile benannt. Die eingeschränkte Objektivität von mündlichen Prüfungen wurde von der Mehrheit der Prüfenden bemängelt. Die Validität der mündlichen Prüfung für die praktische Arbeit mit Patient*innen wurde insbesondere vor dem Hintergrund der komplexen Langzeitbetreuung multimorbider Patient*innen in der Allgemeinmedizin kritisch hinterfragt.

> *„Wir sehen nur das Idealbild, was ein Prüfling, ein Kandidat uns erzählt verbal, was er bei den dargestellten Fällen macht. Wir haben keine Möglichkeit in seinen Alltag, in seine Performance und in die Arzt-Patienten-Kommunikation hineinzugucken" (#1)*

Als strukturelle Nachteile wurden von einigen Prüfeneden der späte Prüfungszeitpunkt mit begrenztem Einfluss auf die Weiterbildung sowie die Unsicherheit als Prüfende und der Aufwand bei Nichtbestehen genannt. Im Kontext der Vor- und Nachteile wurde betont, dass nicht abgebildete Kompetenzen dennoch vorhanden sein können und die Prüfung nur begrenzt geeignet ist, in der verfügbaren Zeit umfassend und vertieft Kompetenzen zu erfassen:

> *„In 30 Minuten kann man aus meiner Sicht gar nicht so viele Ziele haben. Es geht darum, dass die Prüfer sich ein Bild davon machen sollen, ob der/diejenige, der/die geprüft wird, auch in der Lage ist, alleine selbstständig, ohne Aufsicht, allgemeinmedizinisch in eigener Praxis tätig zu sein" (#2)*

Tabelle 3: Vor- und Nachteile aktueller Facharztprüfungsform in der Wahrnehmung von Facharztprüfenden

Vorteile		Nachteile	
organisatorische Vorteile	Niederschwelligkeit	Mangel an Objektivität und Standardisierung	Subjektivität (Prüfendenabhängigkeit, unterschiedliche Maßstäbe)
	Praktische Umsetzbarkeit		Fehlende Standarddurchführung
	Geringer Zeitaufwand		Wunsch nach definierten Kriterien (Erwartungshorizont)
fachliche Qualität	Prüfung durch Fachkolleg*innen	eingeschränkte Validität	Geringe Aussagekraft für praktische Tätigkeit
	Qualitätskontrolle		Geringe Aussagekraft für Komplexität hausärztlichen Handelns
	Prüfung von Haltungen		Schwierigkeit Differenzierung in kurzer Zeit
Vorteile Prüfungssetting	Einzelprüfung	Strukturelle Nachteile	Prüfung erst am Ende der Weiterbildungszeit
	Persönliche Prüfung		Aufwand/Unsicherheit für Prüfende im Falle Nichtbestehens
Gestaltungsspielraum für Prüfende	Freiheit als Prüfende bzgl. Fragen-/ Themenauswahl	Potenzielles Manipulations-/ Ausnutzungsrisiko	Gefahr Ausnutzung des Prüfungsformats durch Prüflinge

Kriterien für Prüfende

Alle Prüfenden benannten, dass sie von den zuständigen Kammern ernannt wurden und ihnen zumindest kein transparentes Verfahren bekannt sei, wie Prüfende ausgewählt würden. In Tabelle 4 sind die von den Prüfenden berichteten Kriterien, die sie für die Auswahl von Prüfenden anlegen würden.
Inwiefern objektive Kriterien die persönliche Eignung von Prüfenden widerspiegeln könnten, wurde von mehreren Prüfenden kritisch hinterfragt. Als mögliches Verfahren wurde von einem Prüfer beispielsweise ein kollegiales Gespräch (Prüfendenprüfung) vorgeschlagen, in dem anhand von Fallbesprechungen und der Interaktion die fachliche und persönliche Eignung von Prüfenden besser eingeschätzt werden könne.

Veränderungsperspektiven

Die Prüfenden zeigten verschiedene Sichtweisen (siehe Tabelle 5), inwiefern es bei der aktuellen Prüfungsform Veränderungsbedarf gibt. Teilweise wurde der Status quo als ausreichend gut bewertet und eine „Verschulung" nach Prüfungskriterien der Universität abgelehnt. Die Mehrheit setzte den Fokus auf die Optimierung und Weiterentwicklung der bestehenden Prüfungsform. Hier wurden insbesondere vorher gemeinsam konsentierte Erwartungshorizonte zur Verbesserung des aktuellen Prüfungsprozesses benannt. Zudem wurde von manchen Prüfenden die Ergänzung der Prüfung sowohl um eine schriftliche als auch praktische Prüfung überlegt. Drei Prüfende äußerten den Wunsch nach internationalem Vorbild ein longitudinales Assessment, um Prozesse und Kompetenzen besser abzubilden. Es wurde von einem Prüfer zudem benannt, inwiefern ein solches Assessment auch auf freiwilliger Basis stattfinden könnte. Der Impact einer Veränderung der Prüfung auf die Weiterbildung wurde unterschiedlich eingeschätzt. Während einige wenig Potential hierin sehen, sahen andere eine deutliche Chance hierin:

> *„Meines Erachtens kann eine Veränderung der Prüfung sehr dazu beitragen, die Qualität und auch die Zufriedenheit während der Weiterbildung zu verbessern. Also wenn breiter geprüft werden würde, dann müssten ja die Kollegen in Weiterbildung nachziehen und anders lernen oder auch andere Ansprüche an ihre Weiterbilder stellen. Also, da muss ja nachgezogen werden und das könnte die Weiterbildung verändern." (#10)*

Generell waren sich die Mehrheit der Prüfenden einig, dass der Fokus auf der Weiterentwicklung und Verbesserung der Weiterbildung und nicht auf einer komplizierten Prüfung.

Effekte der Kompetenzzentren Weiterbildung

Die Hälfte der interviewten Prüfenden weiß von den Prüflingen (z.B. aufgrund aktiver Nachfrage oder Bekanntheit über Seminare im KW), ob sie im Kompetenzzentrum waren und können Effekte der KWs als deutliche Tendenzen verorten.
Diese Prüfenden benannten, dass Teilnehmende des KW einen besseren Prüfungserfolg hätten und in der Heran-

Tabelle 4: Benannte Kriterien für Prüfende im Fach Allgemeinmedizin aus Sicht der Prüfenden

Fachliche Kompetenz und Berufserfahrung	Relevante Erfahrung in hausärztlicher Tätigkeit	*#5: „Ja, Hausärztliche Tätigkeit, aber das ist, denke ich mal, ein implizites Kriterium. Ich kann natürlich nur gut und fair prüfen, wenn ich seit geraumer Zeit tatsächlich hausärztlich tätig bin.“*
	Evidenzbasiertes, leitliniengerechtes Arbeiten	*#10: „Also ganz wichtig ist evidenzbasierte Medizin, Ob die da Leitlinien orientiert arbeiten oder nicht, oder auch mal wissen, wann sie von Leitlinien abweichen können.“*
	Reflektierte haus-ärztliche Arbeitsweise	*#9: „Allgemeinmedizin mehr ist als die Evidenz aus Studien. Kann jemand sein Handeln in Bezug auf die hausärztliche Arbeitsweise reflektieren und auch sich selbst hinterfragen.“*
Prüfungs- und Lehrerfahrung	Vorerfahrung in Konzipierung und Durchführung von Prüfungen	*#6: „also Prüfungserfahrungen in anderen Zusammenhängen. Es gibt ja Prüfungen, also Leute, die zum Beispiel in der MFA-Prüfung tätig sind, die also schon ein bisschen Ahnung haben, was in Prüfungen geschieht und wie sie umzugehen haben mit den Kandidaten.“*
	Allgemeine Lehrerfahrung	*#11: „An zweiter Stelle fände ich es gut, wenn die Leute eine gewisse Lehrerfahrung hätten.“*
	Weiterbildungs-erfahrung/-befugnis	*#9: „eine Weiterbildungsermächtigung, auf jeden Fall, dass man selber auch weitergebildet hat. Vielleicht nicht nur die Weiterbildungsermächtigung, sondern auch die Erfahrung in der Weiterbildung.“*
Prüfenden-schulung/ gemeinsame Standards	Teilnahme an Prüfendenschulung	*#2: „Und dass ich im Anschluss daran eine Prüferschulung durchführen würde, in dem man zum einen mit ihnen gemeinsam Ideen entwickeln lässt, lässt sie auch Falllinien entwickeln und lässt sie dann unter, am besten sogar Video- und Tonaufzeichnungen, eine Prüfung durchführen und bespricht im Anschluss mit ihnen, ob es dort noch Ideen der Verbesserung gibt.“*
	Prüfendenprüfung („Prüfende selbst prüfen“)	*#2: „Ich würde zunächst dahin gehen und würde sagen, dass jeder Prüfer selbst sich einer Prüfung unterziehen muss. Das fände ich ganz spannend, wenn Prüfer geprüft werden, bei denen man kollegiales Gespräch führt. Das muss nicht so ein kleines Prüfungssetting haben, aber dass man die einfach mal einlädt und spricht so eine Stunde mit ihnen. Wie gestalten sie so ihr Leben, wie gestalten sie die Patientenkommunikation, wie gestalten sie die Fälle. Lassen wir uns gerade mal über ein paar Fälle sprechen.“*
	Verpflichtung zur Erstellung von Erwartungshorizonten	*#11: „dass man sich verpflichtet, vorher einen Erwartungshorizont zu erstellen und zu besprechen. Das sind zwei DIN A4 Seiten Aufwand, aber dafür sollte man bereit sein.“*
Persönliche Eignung und soft skills	Teamfähigkeit	*#10: „dann wäre mir wichtig so ein Teamplayer, dass es Menschen sind, die als Teamplayer auch aufgestellt sind.“*
	Persönliche Integrität	*#9: „Und die persönliche Eignung: Die Persönlichkeiten sind halt schwierig zu kategorisieren. Ich sollte auch nett und ausgewogen und empathisch sein. Aber bei der persönlichen Eignung ist es schwierig anzusetzen.“*
	Bereitschaft, Prüfungen abzunehmen	*#1: „Es ist schwierig genug, Leute zu finden, die den Weg und den Sprechstundenausfall auf sich nehmen.“ #7: „An erster Stelle steht die Bereitschaft, es zu tun.“*
Diversität in Prüfungs-komitees	Soziodemographische Charakteristika	*#9: „Dann nutze ich tatsächlich eine Mischung, wen wir gerne haben wollen: weibliche und männliche, jünger und älter in der Zusammensetzung der Komitees.“*
	Erfahrungshintergründe	*#4: „Das finde ich etwas ganz Wichtiges, was ja eigentlich nicht abgebildet wird, wie man den Spagat zwischen wirtschaftlichem Handeln und Führung von Personal und Leitlinie und Patientenbilder zusammenbringt. Das müsste ein Prüfer selber in seiner Vitae und in seiner praktischen Erfahrung abbilden.“*
	Selbstständigkeit/ Anstellung	*#3: „In unserem derzeitigen System ist die Tätigkeit als selbstständig niedergelassener wichtig, um beurteilen zu können, wie Antworten ausfallen, die normal von den eigenen Einstellungen abweichen. Wirtschaftsgesichtspunkte, Workflow-Gesichtspunkte.“*

Tabelle 5: Perspektiven von Facharztprüfenden auf Veränderung der aktuellen Prüfungsform

Beibehaltung status quo	Keine Veränderung, Zufriedenheit mit Status Quo	*#4: „Eine Veränderung mit regelmäßigen Assessments finde ich bürokratisch und dumm ehrlich gesagt, weil jeder seine eigene Weiterbildung stricken muss. Und letztendlich ist die Allgemeinmedizin hinterher so bunt, wie die Gesellschaft verschiedene Ärzte und Typen braucht. Ich finde es so, wie es ist, gut.“*
	Prüfung nicht mit Prüfungskriterien „verschulen“	*#4: „Ne, ich wäre sehr dagegen, das noch weiter zu verschulen.“*
Verbesserung Prüfungs-organisation	Austausch zwischen den Prüfenden vor der Prüfung	*#8: „Kurzen Austausch zwischen den beiden Prüfern, zwei, drei Tage vorher. Es ist so, dass die Prüflinge erst am Tag der Prüfung erfahren, wer sie prüft. Wir wissen das schon, aber wir kriegen auch nicht mitgeteilt, mit wem wir prüfen. Das fände ich gut, wenn wir uns austauschen könnten.“*
	Wunsch nach internen Kriterien für die Prüfung/ Erwartungshorizonte	*#9: „Also ich würde ein paar interne Kriterien anlegen wollen, was so auf jeden Fall geprüft werden soll, welche Bereiche und was so eine Basis zum Bestehen ist und was so darüber hinausgeht. Also man hat ja so Prüflinge, wo man sagt, er hat zumindest niemand umgebracht. Und dann gibt es die Leute, wo du sagst, hey, gib mir Deine Adresse, ich möchte Dich sofort anstellen. Also so ein paar Basiskriterien, das reicht so zu bestehen.“*
Weiterentwicklung des bestehenden Prüfungsformats	Aktuelle Prüfungsform mit diversifizierter Prüfungsmethodik	*#10: „Die Verwendung von mehr echtem Material, sagen wir EKGs, Bilder, Sonographien, Untersuchungstechniken, also sei es Phantome oder Simulationspatienten. Das hätte ich gern. Und natürlich Video. Sprich die Einreichung von Videodokumenten aus der Sprechstunde.“*
	Ergänzung um eine schriftliche Prüfung	*#1: „Ich denke, dass es gut wäre, ein schriftliches Bein zu haben, um den Wissensstandard zu gewährleisten. Beispiel USA Board Certification.“*
	Ergänzung um eine praktische Prüfung	*#1: „Ich denke, dass es gut wäre, mehr Praxis zu haben, OSCE-mäßig.“*
Alternative Perspektiven auf Prüfungsform und Weiterbildung	Wunsch nach Prüfungsformat, das Prozesse und Kompetenzen abbildet	*#7: „Also ich würde mir wirklich wünschen, dass das Prüfungsformat an sich überdacht wird. Also von diesem Gespräch hin zu irgendwas anderem, was Prozesse abbildet und Kompetenzen abbildet.“*
	Wunsch nach longitudinalem Assessment	*#2: „Ich würde mir wünschen, dass wir alle sechs Monate mit den Kollegen ein Gespräch führen, Fähigkeiten üben, Fähigkeiten, die sie gelernt haben, überprüfen. Dass wir wirklich über diese fünf Jahre hinweg regelmäßig bestimmte Kenntnisse abfragen, beziehungsweise Tätigkeiten prüfen. Und völlig klar ist mir, dass das ein Riesenaufwand ist und ein Prozess, der sich nicht von heute auf morgen umsetzen lässt, der meines Erachtens aber sehr dazu beitragen kann, die Qualität und auch die Zufriedenheit während der Weiterbildung zu verbessern.“*
	Fokus auf Qualität der Weiterbildung statt Fokus auf Prüfung	*#6: „Und ich würde mehr gucken, danach die Weiterbildung qualitativ zu steigern. Nicht die Prüfung schadvoll zu machen.“*

gehensweise an die Fragen eine bessere Methodenkompetenz und Struktur zeigen würden. Zudem schreiben sie den Teilnehmenden des KW mehr praktische Handlungskompetenz zu. Gleichzeitig wird die Abgrenzung zu anderen Faktoren als schwierig erachtet. Beispielsweise wurde die diversifizierte Weiterbildung mit verschiedenen ambulanten Rotationen als deutlicher Prädiktor für eine bessere Performance bei der Prüfung benannt. Die Prüfenden stellten klar, dass Nicht-Teilnehmende am KW ebenfalls sehr gute Prüfungen ablegen können. Ein von ihnen wahrgenommener Trend zeigt, dass die Mehrheit der Nichtbestehenden nicht am KW Weiterbildungen besucht habe:

> *„Also was man sagen kann, ist die, die am Kompetenzzentrum angebunden sind, tatsächlich recht gut in den Prüfungen abschneiden. Es bedeutet nicht, dass jemand, der nicht im Kompetenzzentrum war, nicht gut sein kann, aber die, die nicht gut waren, die waren in der Regel nicht am Kompetenzzentrum angebunden. So könnte man das sagen, genau“ (#11)*

Diskussion

Die qualitative Studie zeigt den Status quo der Facharztprüfung mit ihren Vor- und Nachteilen sowie Weiterentwicklungspotentialen und Effekten der Kompetenzzentren Weiterbildung auf. Es konnten die Perspektiven der Prüfenden auf die Prüfungsform, auf Veränderungsmöglichkeiten sowie Qualifikation von Prüfenden explorativ erfasst werden.

Der demographische Wandel der Ärzt*innen [22] macht auch bei den Prüfenden nicht halt und vor dem Hintergrund der steigenden Facharztanerkennungen in der Allgemeinmedizin lässt sich ein Bedarf für die Rekrutierung und Qualifizierung neuer Prüfender ableiten. Bezüglich der Motivation für die Prüfungstätigkeit wurden vor allem intrinsische und fachbezogene Motive benannt. Ausreichende finanzielle Anreize zur Kompensation von Sprechstundenausfällen könnte die Rekrutierung weiterer Prüfender erleichtern.

Es besteht eine Ambiguität in der Haltung zur Facharztprüfung durch die Prüfenden. Die Prüfenden waren sich einig, dass sie in der Prüfung den Mindeststandard einer Facharztqualifikation eruieren können, wenngleich sie deutliche Einschränkungen in Hinblick auf Objektivität, Reliabilität, Validität und Kontextsensibilität sahen. Dies deckt sich mit internationalen Untersuchungen zu mündlichen Prüfungen. Memon et al. [23] zeigen in ihrem Review auf, dass in vielen Ländern mündliche Prüfungen in den postgraduellen Weiterbildungen traditionell etabliert sind, wenngleich manche Länder aufgrund der eingeschränkten testtheoretischen Güte andere Prüfungsformate eingeführt haben. In der Schweiz wird beispielsweise eine schriftliche Prüfung mit 120-MC-Fragen durchgeführt, was mit einer erhöhten Objektivität einhergeht [24]. Burch et al. [25] sehen die Evidenzbasierung der Prüfungsformate in Hinblick auf internationale Vergleichbarkeit bezüglich ihrer Aussagekraft kritisch. Entsprechend der Miller-Pyramide [15] können in mündlichen Prüfungen vor allem kognitive und Methodenkompetenzen („weiß es“ (*knowledge*), „weiß wie“ (*competence*)) geprüft werden, wohingegen Handlungskompetenzen („zeigt es“ (*performance*), „macht es“ (*action*)) nicht in mündlichen Prüfungen adäquat abgebildet werden können. Die Überprüfung des eigenständigen professionellen Handelns stellt das genuine Ziel der Facharztprüfung als Abschluss von Aus- und Weiterbildung dar.

Bezüglich Veränderungsimplikationen bestand unter den Prüfenden eine deutliche Meinungsheterogenität. Es lässt sich unter den Prüfenden wenig Anreiz für Veränderung eruieren. Fink et al. [17] konnten unter Ärzt*innen in Weiterbildung in einer Umfrage ebenfalls wenig Interesse an einer Veränderung der aktuellen Prüfungsform feststellen. Derzeit gibt es keine Hinweise auf eine Bottom-Up-Initiative zur Veränderung der aktuellen Prüfung, weder von Prüfenden noch von Ärzt*innen in Weiterbildung. Gleichzeitig wurden kurzfristige und leicht implementierbare Optimierungen der aktuellen Prüfungsform mit Prüfendenschulungen, regelmäßigem Austausch (Qualitätszirkel für Prüfende) und gemeinsam konsentierten Erwartungshorizonten benannt. Dies könnte die Durchführungsobjektivität, die Auswertungsobjektivität und die Interpretationsobjektivität erhöhen. Beckers et al. [26] zeigten in einer Analyse von Gedächtnisprüfungsprotokollen der Facharztprüfung Innere Medizin, dass die Varianz der Prüfungsinhalte und Kompetenzstufen relativ groß ist und empfehlen Blueprints wie in den USA für gleiche Prüfungsbedingungen für alle Kandidat*innen. In verschiedenen europäischen Ländern durchlaufen Prüfende bereits vor ihrem ersten Einsatz eine Prüfendenschulung. In Südbaden durchgeführte Prüfungsschulungen [28] erfuhren eine große Akzeptanz durch die teilnehmenden Prüfenden.

Diese sahen die Notwendigkeit, Minimalstandards für Prüfungen festzulegen und Anforderungsniveau und Bewertungskriterien zu vereinheitlichen, um Validität und Reliabilität der Prüfungen zu verbessern.

Auf Basis der von den Prüfenden in der Studie benannten Qualifikationskriterien lässt sich ein Absolvierendenprofil von Prüfendenschulungen ableiten, das die Grundlage für die Weiterentwicklung entsprechender Lehr- und Lernziele für Prüfendenschulungen bilden kann. Die Schulungen sollten sowohl die Durchführungskompetenz und das Wissen über unterschiedliche Prüfungsformate vermitteln, als auch Raum für eine reflexive Diskussion zwischen praktischer Umsetzbarkeit und prüfungsdidaktischem Anspruch bieten. Für eine Restrukturierung der Prüfung, die machbar gelebt wird, ist die konsequente Integration von Prüfenden notwendig.

Analog zur Verpflichtung für die Neubeantragung einer Weiterbildungsbefugnis einen Train-the-Trainer-Kurs zu besuchen [29] könnte die Verpflichtung einer kompakten Prüfendenschulung für neue Prüfende empfohlen werden. Solche Schulungen sind beispielsweise in Dänemark, den Niederlanden und Großbritannien üblich [27]. Um Unsicherheiten und Verwaltungsaufwand bei Nichtbestehen für die Prüfenden zu reduzieren, könnten sie gezielt zu Abläufen geschult werden, während vorab erstellte Erwartungshorizonte eine belastbare und nachvollziehbare Entscheidungsgrundlage bieten. In der Musterweiterbildungsordnung 2018 sind ausschließlich formale Aspekte der Prüfung beschrieben [1]. Vor dem Hintergrund der Transition hin zur Kompetenzorientierung der neuen Musterweiterbildungsordnung wäre eine entsprechende kompetenzorientierte Ausrichtung des Prüfungsformats zielführend. Vorschläge zu strukturierten mündlichen Prüfungen für ein qualitativ höheres und einheitliches Niveau existieren [30]. Für eine kontinuierliche Sicherstellung des Weiterbildungsprozesses sind formative Assessments (*assessment for learning*) besser geeignet als summative Prüfungen (*assessment of learning*). In der psychotherapeutischen Ausbildung in Deutschland gibt es beispielsweise eine Zwischenprüfung und eine Abschlussprüfung. Letztere besteht aus einem schriftlichen MC-Teil sowie einer mündlich-praktischen Fallprüfung [https://www.gesetze-im-internet.de/psychthappro/BJNR044800020.html]. International haben sich in der Allgemeinmedizin formative Assessments beispielsweise

in Dänemark und den Niederlanden [31], [32] bewährt und es wird auf eine summative Abschlussprüfung verzichtet. In Frankreich dokumentieren Weiterzubildende ihre Kompetenzentwicklung in einem Portfolio, das von Mentor*innen (référent pédagogique) kontrolliert wird [33]. Khan et al. [34] zeigen in ihrem Review ebenfalls Vorteile von formativem Assessment für die Verbesserung der individuellen Lernkurve in der Urologie auf. Kompetenzen sollten kontinuierlich durch formativ-praktische Assessments wie Mini Clinical Examination (Mini-CEX), Direct Observation of Procedural Skills (DOPS) oder Portfolios erfasst und rückgemeldet werden [35]. Hier zeigen sich in den letzten Jahren entrustable professional activities (EPAs) vielversprechend, um die Handlungskompetenzen in der Miller-Pyramide aussagekräftig abzubilden [36], wenngleich eine Studie aus den Niederlanden in EPAs einen größeren Vorteil als Lerninstrument und weniger als Prüfungsinstrument in der Allgemeinmedizin sieht [37]. Thiessen et al. [38] entwickelten ein Best-Practice-Modell für die Facharztprüfung mit Triangulierung durch verschiedene Methoden, Blueprinting, Festlegung von Fragestellungen und Erwartungshorizont, Qualitätskontrolle und nachgelagerten Lerneffekten für Prüflinge. Der erhöhte Prüfungsaufwand geht entsprechend mit höheren Kosten und mehr Personalaufwand einher [27]. Es zeigt sich eine Ambivalenz zwischen dem Bewusstsein für die Schwächen der aktuellen Prüfungsform und der Motivation für eine grundlegende Reform. Aus prüfungstheoretischer Perspektive erscheint eine Weiterentwicklung jedoch sinnvoll, um ein besseres Alignment zwischen Kompetenzzielen und Prüfungsformaten zu erreichen. Vor dem Hintergrund der Bedeutung von Prüfungen sowohl für Lernprozesse als auch für die Qualitätssicherung in der medizinischen Versorgung ist eine solche Anpassung aus prüfungsdidaktischer Sicht erforderlich. Mehrere Prüfende betonten zugleich Implementierungsaspekte, insbesondere die begrenzte Verfügbarkeit von Prüfenden und die praktische Umsetzbarkeit.

Ein longitudinales formatives Assessment (Portfolio-basiert mit Logbuch) mit praktischen Prüfungsanteilen könnte in Deutschland ebenfalls die Qualität und Effizienz in der Weiterbildung stärken. Derzeit wird dies teilweise über Weiterbildungsbefugte und das Logbuch erfasst, eine unabhängige Prüfungsinstitution zur triangulierenden Kontrolle wäre jedoch vorteilhaft. Da viele Prüfungskandidat*innen über Verlaufs- oder Quereinstieg [39] nicht primär in der Allgemeinmedizin sozialisiert sind, könnte ein formatives Assessment frühzeitig Feedback zu ihrer Neuorientierung und den Lernzielen für die verbleibende Weiterbildungszeit geben. Nach internationalem Vorbild gibt es auch Progress Tests [40] in der Weiterbildung, die den Ärzt*innen in Weiterbildung Rückmeldung über ihren Lernfortschritt und zukünftige Lernziele geben. Es wäre also auch eine freiwillige Verpflichtung zu formativen Assessments zur Rückmeldung über den eigenen Lernfortschritt denkbar.

Ein Teil der Prüfenden konnte den Teilnehmenden der KW eine bessere Prüfungsperformance und eine geringere Nichtbestehens-Quote zuschreiben. Diese Tendenzen lassen sich als potenzielle Effekte der Kompetenzzentren Weiterbildung (KW) interpretieren. Gleichzeitig ist ein Selektionseffekt von besonders motivierten Ärzt*innen in Weiterbildung und weitere Faktoren (z.B. bessere Vernetzung) zu diskutieren. Bemerkenswert ist insbesondere die Feststellung, dass zumeist Prüflinge mit Nichtbestehen in der Prüfung keine Weiterbildungen im KW besuchten.

Stärken und Limitationen

Es handelt sich um eine qualitative Erhebung von Prüfenden mit sowohl langjähriger Prüfungserfahrung als auch wenig Prüfungserfahrung, was die Aussagekraft der Studie untermauert. Gleichzeitig sind verschiedene Limitationen aufzuführen. Die qualitative Befragung hat keinen Anspruch auf Vollständigkeit, da nicht Prüfende aus allen Bundesländern befragt wurden. Die Gesamtzahl der Prüfenden für Allgemeinmedizin in Deutschland ist derzeit nicht bekannt. Die Stichprobe umfasste sowohl wenig erfahrene, als auch überwiegend erfahrene Prüfende und dürfte die Zusammensetzung von Prüfungsgremien plausibel abbilden. Es ist ein möglicher Selektionsbias zu diskutieren, da die Rekrutierung teilweise über bestehende Netzwerke erfolgte und sich vor allem schon bekannte Prüfende zurückmeldeten. Es ist anzunehmen, dass möglicherweise universitär vernetzte Prüfende neuen Prüfungsformaten offener gegenüber eingestellt sind. Zudem ist im Kontext der Interviews eine Beeinflussung durch soziale Erwünschtheit zu diskutieren.

Die berichteten Effekte der Kompetenzzentren Weiterbildung sind als explorativ zu interpretieren, da nicht alle Prüfenden die Teilnahme der Prüflinge entsprechend einordnen konnten und somit wahrscheinlich nicht alle relevanten Perspektiven erfasst wurden.

In weiteren Studien sollte im Mixed-Methods-Design die Einschätzung der Umsetzbarkeit und des Mehrwehrts für die Weiterbildung zur Veränderung der Facharztprüfung sowie die Effekte der Kompetenzzentren weiter untersucht werden.

Schlussfolgerungen

Die Prüfenden zeigen ein ausgeprägtes Bewusstsein für die Schwächen der aktuellen Prüfungsform und reflektieren kritisch deren Objektivität, Reliabilität und Validität im Hinblick auf die Komplexität allgemeinmedizinischer Tätigkeit. Hinsichtlich zielführender Veränderungen besteht Meinungsvielfalt, jedoch wird die Einschätzung des Mindeststandards der Facharztqualifikation in der aktuellen Form aus der Perspektive der Prüfenden als möglich erachtet. Aus prüfungsdidaktischer Sicht sollte die Zieldefinition der Facharztprüfung geschärft, das Alignment von Kompetenzzielen der Musterweiterbildungsordnung und Prüfungsformat verbessert und zur Qualitätssicherung in der Weiterbildung ein longitudinales Assessment mit praktischen Prüfungsanteilen erwogen werden. Für eine zeitnahe Veränderung erscheinen flächendeckende Prü-

fendenschulungen, regelmäßiger Austausch unter den Prüfenden und konsentierte Erwartungshorizonte geeignet.

Anmerkungen

Ethische Genehmigung und Einwilligung zur Teilnahme

Die Studie entsprach der Deklaration von Helsinki. Alle Teilnehmer haben ihre schriftliche Einwilligung zur Teilnahme an der Studie und zur pseudonymisierten Verwendung ihrer Daten erteilt. Die Studien wurden von der Ethikkommission der Medizinischen Universität Heidelberg, Alte Glockengießerei 11/169115 (S711/2023), genehmigt. Die Studien wurden in Übereinstimmung mit den lokalen Rechtsvorschriften und institutionellen Anforderungen durchgeführt. Die Teilnehmer haben ihre schriftliche Einwilligung nach Aufklärung zur Teilnahme an dieser Studie erteilt.

Verfügbarkeit von Daten und Materialien

Die in diesem Artikel vorgestellten Datensätze sind nicht frei zugänglich, da die Teilnehmer gemäß den Anforderungen des europäischen Datenschutzrechts über die Datenschutzrichtlinien informiert wurden und zugestimmt haben, dass die pseudonymisierten Daten dem Projektteam zur Verfügung stehen und am Studienzentrum gespeichert werden. Anfragen bezüglich des Zugriffs auf die Datensätze sind an Simon Schwill (simon.schwill@med.uni-heidelberg.de) zu richten.

Förderung

Die Autor*innen erklären, dass für die Forschung, die Erstellung und/oder die Veröffentlichung dieses Artikels finanzielle Unterstützung gewährt wurde. Der KWBW Verbundweiterbildungplus® wird durch öffentliche Mittel gemäß § 75a des Sozialgesetzbuchs V, Anhang, gefördert.

Beitrag der Autor*innen

ACD war an der Ideenfindung, der Konzeption und Methodik sowie an der Erhebung, Analyse und Interpretation der Daten und der Erstellung des ersten Entwurfs beteiligt. CB war an der Erhebung, Analyse und Interpretation der Daten beteiligt. MR war an der Methodik sowie an der Analyse der Daten beteiligt. AA war an der Auswertung der Daten und der Überprüfung des Manuskripts beteiligt. SS war an der Idee und der Finanzierung der Studie, an der Konzeption und Methodik sowie an der Analyse und Auswertung der Daten und der Erstellung des ersten Entwurfs beteiligt.

ORCIDs der Autor*innen

- Andreas Christian Dreher: [0009-0003-6688-8807]
- Christine Becker: [0009-0009-4802-5761]
- Marco Roos: [0000-0003-1596-5908]
- Attila Altiner: [0000-0002-2429-933X]
- Simon Schwill: [0000-0002-0954-2194]

Danksagung

Wir möchten uns bei allen teilnehmenden Hausarztprüfer*innen, allen KWBW-Praxen und Partner*innen für ihre Teilnahme bedanken. Die Studie wurde auf der Jahrestagung der DACH-Gesellschaft für medizinische Ausbildung in den Gesundheitsberufen 2025 in Düsseldorf vorgestellt.

Interessenkonflikt

ACD, CB und SS erklären, dass sie in die Organisation des Weiterbildungsprogramms KWBW eingebunden sind. Alle Autoren erklären, dass bei ihnen keine Interessenkonflikte bestehen.

Anhänge

Verfügbar unter https://doi.org/10.3205/zma001872

1. Anhang_1.pdf (419 KB)
   COREQ (COnsolidated criteria for REporting Qualitative research) Checklist (nur auf Englisch)
2. Anhang_2.pdf (332 KB)
   Facharztprüfung/Effekte KW Facharztprüfer*innen
3. Anhang_3.pdf (152 KB)
   Ankerzitate

Literatur

1. Bundesärztekammer. Postgraduate training curriculum: Muster-Weiterbildungsordnung 2018 in der Fassung vom 26.06.2021. Berlin: Bundesärztekammer; 2021.
2. Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. Med Teach. 2007;29(7):642-647. DOI: 10.1080/01421590701746983
3. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrot V, Roberts T. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):206-214. DOI: 10.3109/0142159X.2011.551559
4. Bundesärztekammer. Ärztestatistik zum 31. Dezember 2024. Berlin: Bundesärztekammer; 2024. Zugänglich unter/available from: https://www.bundesaerztekammer.de/fileadmin/user_upload/BAEK/Ueber_uns/Statisik/AErztestatistik_2024.pdf

5. Landesärztekammer Baden-Württemberg. Neue Ärztestatistik für Baden-Württemberg. Stuttgart: Ärztekammer Baden-Württemberg; 2024. Zugänglich unter/available from: https://www.aerztekammer-bw.de/neue-aerztestatistik-fuer-baden-wuerttemberg-f14fcf0d7b2267dc

6. Dißmann P. Europäische Facharztprüfung für Notfallmedizin. Notfall Rettungsmed. 2014;17:697-700. DOI: 10.1007/s10049-014-1937-5

7. Schwill S, Magez J, Joos S, Steinhäuser J, Ledig T, Rubik A, Niebling W, Szecsenyi J, Flum E. New paths in post-graduate medical training in general practice - 8 years of experience with the pilot project Verbundweiterbildungplus Baden-Württemberg. GMS J Med Educ. 2017;34(5):Doc62. DOI: 10.3205/zma001139

8. Elmer A, Grifka J. Vergleich von Prüfungsmethoden in der medizinischen Ausbildung. Gesundheitswesen (Suppl Med. Ausbild). 1998;15(Suppl1):14-17. Zugänglich unter/available from: https://gesellschaft-medizinische-ausbildung.org/files/ZMA-Archiv/1998/1/Elmer_A,_Grifka_J.pdf

9. Artino AR Jr, Durning SJ, Creel AH. Reliability and validity in educational measurement. Acad Med. 2010;85(9):1545. DOI: 10.1097/ACM.0b013e3181edface

10. Wormald BW, Schoeman S, Somasunderam A, Penn M. Assessment drives learning: an unavoidable truth? Anat Sci Educ. 2009;2(5):199-204. DOI: 10.1002/ase.102

11. van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. Med Educ. 2005;39(3):309-317. DOI: 10.1111/j.13652929.2005.02094.x

12. Yazbeck Karam V, Park YS, Tekian A, Youssef N. Evaluating the validity evidence of an OSCE: results from a new medical school. BMC Med Educ. 2018;18(1):313. DOI: 10.1186/s12909-018-1421-x

13. Pell G, Fuller R, Homer M, Roberts T; International Association for Medical Education. How to measure the quality of the OSCE: a review of metrics – AMEE Guide No. 49. Med Teach. 2010;32(10):802-811. DOI: 10.3109/0142159X.2010.507716

14. ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? Acad Med. 2007;82(6):542-547. DOI: 10.1097/ACM.0b013e31805559C7

15. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9):S63-S67. DOI: 10.1097/00001888-199009000-00045

16. Biggs J. Enhancing teaching through constructive alignment. High Educ. 1996;32(3):347-364. DOI: 10.1007/BF00138871

17. Fink M, Roos M, Schwill S, Steinhäuser J. Facharztprüfung Allgemeinmedizin: Quo vadis? ZFA (Stuttgart). 2025;101:316-321. DOI: 10.1007/s44266-025-00377-y

18. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349-357. DOI: 10.1093/intqhc/mzm042

19. Kern DE, Thomas PA, Hughes MT. Curriculum development for medical education: a six-step approach. 2nd ed. Baltimore: Johns Hopkins University Press; 2009.

20. Van Someren M. The think aloud method. London: Academic Press; 1994.

21. Kuckartz U, Rädiker S. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. 5. überarbeitete Auflage. Weinheim: Beltz Juventa; 2022.

22. Bundesärztekammer. Arztzahlstatistik: Jede/r vierte berufstätige Ärztin/Arzt ist 60 Jahre und älter. Berlin: Bundesärztekammer; 2023. Zugänglich unter/available from: https://www.bundesaerztekammer.de/presse/aktuelles/detail/arztzahlstatistik-zeigt-jede-r-vierte-berufstaetige-aerztin-arzt-ist-60-jahre-und-aelter

23. Memon MA, Joughin GR, Memon B. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. Adv Health Sci Educ Theory Pract. 2010;15(2):277-289. DOI: 10.1007/s10459-008-9111-9

24. Schweizerisches Institut für ärztliche Weiter- und Fortbildung (SIWF). Fachärztin oder Facharzt für Allgemeine Innere Medizin. Weiterbildungsprogramm. Bern: Schweizerisches Institut für ärztliche Weiter- und Fortbildung; 2022. Zugänglich unter/available from: https://www.siwf.ch/files/pdf21/aim_version_internet_d.pdf

25. Burch VC, Norman GR, Schmidt HG, van der Vleuten CP. Are specialist certification examinations a reliable measure of physician competence? Adv Health Sci Educ Theory Pract. 2008;13(4):521-533. DOI: 10.1007/s10459-007-9063-5

26. Beckers MA, Braun LT, Epstein N, Fischer MR, Schmidmaier R. Die Facharztprüfung Innere Medizin: Eine qualitative und quantitative Analyse von Gedächtnisprotokollen. Dtsch Med Wochenschr. 2020;145(2):e10-e17. DOI: 10.1055/a-0970-6348

27. Flum E, Maagaard R, Godycki-Cwirko M, Scarborogh N, Scherpbier N, Ledig T, Roos M, Steinhäuser J. Assessing family medicine trainees – what can we learn from the European neighbours? GMS Z Med Ausbild. 2015;32(2):Doc21. DOI: 10.3205/zma000963

28. Streitlein-Böhme I, Niebling W, Fabry G, Böhme K. Three years of experience with a workshop for medical specialty examiners in South Baden: A project report and initial evaluation results. GMS J Med Educ. 2017;34(5):Doc55. DOI: 10.3205/zma001132

29. Bundesärztekammer. Beschlussprotokoll des 128. Deutschen Ärztetags. Berlin: Bundesärztekammer; 2024. Zugänglich unter/available from: https://www.bundesaerztekammer.de/fileadmin/user_upload/BAEK/Aerztetag/128.DAET/2024-05-10_Beschlussprotokoll_neu.pdf

30. Glaab T, Böhme K, Griesinger J, Forster A, Streitlein-Böhme I. Die mündliche Facharztprüfung: Ist das aktuelle Format noch zeitgemäß? Dtsch Med Wochenschr. 2021;146(18):1226-1229. DOI: 10.1055/a-1506-1280

31. IGES Institut. Untersuchung zur Förderung der Weiterbildung Allgemeinmedizin im internationalen Vergleich. Berlin: IGES; 2023. Zugänglich unter/available from: https://www.iges.com/ergebnisse/projekte/2023/weiterbildung-allgemeinmedizin/index_ger.html

32. Miani C, Hinrichs S, Pitchforth E, Bienkowska-Gibbs T, Disbeschl S, Roland M, Nolte E. Best Practice: Medizinische Aus- und Weiterbildung aus internationaler Perspektive. Cambridge (UK): Rand Europe; 2015. Zugänglich unter/available from: https://www.rand.org/content/dam/rand/pubs/research_reports/RR600/RR622z1/RAND_RR622z1.pdf

33. Held LA. Allgemeinmedizin in Studium und Weiterbildung in Frankreich. ZFA (Stuttgart). 2021;2:78-83. DOI: 10.3238/zfa.2021.0078-0083

34. Khan RN, Siddiqui NA. The Use of Formative Assessment in Postgraduate Urology Training: A Systematic Review. Cureus. 2022;14(7):e27162. DOI: 10.7759/cureus.27162

35. Driessen E, Scheele F. What is wrong with assessment in postgraduate training? Lessons from clinical practice and educational research. Med Teach. 2013;35(7):569-574. DOI: 10.3109/0142159X.2013.798403

36. O'Dowd E, Lydon S, O'Connor P, Madden C, Byrne D. A systematic review of 7 years of research on entrustable professional activities in graduate medical education, 2011-2018. Med Educ. 2019;53(3):234-249. DOI: 10.1111/medu.13792

37. Bonnie LHA, Visser MRM, Bont J, Kramer AWM, van Dijk N. Trainers' and trainees' expectations of entrustable professional activities (EPAs) in a primary care training programme. Educ Prim Care. 2019;30(1):13-21. DOI: 10.1080/14739879.2018.1532773

38. Thiessen N, Fischer MR, Huwendiek S. Assessment methods in medical specialist assessments in the DACH region - overview, critical examination and recommendations for further development. GMS J Med Educ. 2019;36(6):Doc78. DOI: 10.3205/zma001286

39. Dreher AC, Ko J, Becker C, Förster C, Jähnig T, Bischoff M, Stengel S, Schwill S. Analyse der Zielgruppen in der Weiterbildung Allgemeinmedizin: Ergebnisse einer Kohortenstudie im Kompetenzzentrum Weiterbildung Baden-Württemberg (KWBW) [Analyzing target audience in family medicine specialty training: Results of a cohort study at the Competence Center for Postgraduate Medical Training]. Z Evid Fortbild Qual Gesundhwes. 2025:197;55-61. DOI: 10.1016/j.zefq.2025.06.004

40. Dijksterhuis MG, Scheele F, Schuwirth LW, Essed GG, Nijhuis JG, Braat DD. Progress testing in postgraduate medical education. Med Teach. 2009;31(10):e464-e468. DOI: 10.3109/01421590902849545

Korrespondenzadresse:
PD Dr. med. Simon Schwill, MME
Universitätsklinikum Heidelberg, Abteilung für Allgemeinmedizin und Versorgungsforschung, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Deutschland, Tel.: +49 (0)6221/56-39379
simon.schwill@med.uni-heidelberg.de

Bitte zitieren als
Dreher AC, Becker C, Roos M, Altiner A, Schwill S. The general practice specialist examination – examiners' perspectives on the status quo and future development: An exploratory qualitative study. GMS J Med Educ. 2026;43(6):Doc78.
DOI: 10.3205/zma001872, URN: urn:nbn:de:0183-zma0018723

Artikel online frei zugänglich unter
https://doi.org/10.3205/zma001872

Eingereicht: *29.10.2025*
Überarbeitet: *30.03.2026*
Angenommen: *30.04.2026*
Veröffentlicht: *15.09.2026*

Copyright
©2026 Dreher 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/.