Aus dem Zentrum Allgemeinmedizin
der Medizinischen Fakultät
der Universität des Saarlandes, Homburg/Saar
Direktor: Prof. Dr. med. Johannes Jäger, MME

Competency-based under- and postgraduate medical education in general practice – an evaluation of tools and outcomes

Dissertation zur Erlangung des Grades eines Doktors der Medizin
der Medizinischen Fakultät
der UNIVERSITÄT DES SAARLANDES

2023

vorgelegt von:
Helene Elisabeth Junge
geb. am: 27.10.1994 in Berlin

Tag der Promotion: 06.09.2023
Dekan: Prof. Dr. med. M.D. Menger
1. Berichterstatter: Prof. Dr. med. J. Jäger
2. Berichterstatter: Prof. Dr. med. S. Wagenpfeil

Table of contents

1 SUMMARY 5

1.1 Deutsche Zusammenfassung 5

1.2 English abstract 7

2 INTRODUCTION 9

2.1 The principle of competency-based medical education in general practice 9

2.2 Intended learning outcomes for general practice specialty training 10

2.3 The role of assessment in competency-based medical education 12

2.4 The general practice blended-learning curriculum at Saarland University 13

2.5 Research question 14

3 METHODS 16

3.1 Ethical considerations 16

3.2 What competencies do European general practice trainees value the most? A prioritisation exercise using a modified Delphi approach 16
3.2.1 Town Hall meeting 17
3.2.2 First Delphi round 18
3.2.3 Second Delphi round 18
3.2.4 Data analysis 19

3.3 The use of digital learning analytics in a blended learning GP curriculum at Saarland University 19
3.3.1 Data analysis 20

4 RESULTS 21

4.1 What competencies do European general practice trainees value the most? A prioritisation exercise using a modified Delphi approach 21
4.1.1 Sociodemographic data 21
4.1.2 Town Hall meeting 22
4.1.3 First Delphi round 24
4.1.4 Second Delphi round 26

4.2 The use of digital learning analytics in a blended learning GP curriculum at Saarland University 28
4.2.1 Sociodemographic data 28
4.2.2 Digital learning behaviour 28
4.2.3 Formative assessment performance 29
4.2.4 Formative assessment performance and academic success 30

5 DISCUSSION 32

5.1 The use of digital learning analytics in a blended learning GP curriculum at Saarland University 32

5.2 What competencies do European general practice trainees value the most? A prioritisation exercise using a modified Delphi approach 33

5.3 Results in the context of literature 35

5.4 Limitations 38

5.4.1 What competencies do European general practice trainees value the most? A prioritisation exercise using a modified Delphi approach 38

5.4.2 The use of digital learning analytics in a blended learning GP curriculum at Saarland University 39

5.5 Conclusion 40

6 REFERENCES 41

7 ABBREVIATIONS 48

8 LIST OF TABLES 49

9 LIST OF FIGURES 51

10 PUBLICATIONS 52

11 APPENDIX 53

12 CV 55

13 ACKNOWLEDGEMENTS 56

1 Summary

1.1 Deutsche Zusammenfassung

Einleitung

In den kommenden Jahren müssen in Deutschland und Europa zahlreiche neue Hausärzte und Hausärztinnen ausgebildet werden, um eine drohende allgemeinmedizinische Unterversorgung zu verhindern. Um die allgemeinmedizinische Ausbildung interessant und qualitativ hochwertig zu gestalten, bietet sich das Konzept der kompetenzbasierten medizinischen Ausbildung an, welches in der Medizindidaktik derzeit hoch relevant ist. Konzepte der kompetenzbasierten Lehre sind unter anderem der Entwurf von konkreten Lernzielen, regelmäßiges Feedback an die Lernenden sowie die stärkere Einbeziehung von Lernenden in die Curriculumsplanung. Im Rahmen zweier Fragestellungen untersucht die vorliegende Dissertation die Implementation von kompetenzbasierter Lehre in der Allgemeinmedizin. Hierzu wurde eine europaweite, strukturierte Kompetenzpriorisierung durchgeführt, um zu ermitteln, welche Kompetenzen von Ärzten und Ärztinnen in der Weiterbildung Allgemeinmedizin als am wichtigsten angesehen werden. In einer seperaten, quantitative Studie wurde der Zusammenhang zwischen den Ergebnissen von formativem und summativem Assessment innerhalb eines blended-learning Curriculums im Medizinstudium untersucht.

Methodik

Zur Untersuchung der Kompetenzprioritäten in der allgemeinmedizinischen Weiterbildung wurde eine strukturierte qualitative Studie nach der modifizierten Delphi-Methode durchgeführt. Zunächst wurden im Rahmen eines internationalen Town Hall Meetings mit jungen Allgemeinmedizinern und Allgemeinmedizinerinnen auf dem Vasco da Gama Movement (heute: European Young Family Doctors Movement) Forum (Edinburgh, Januar 2022) Vorschläge für wichtige Kompetenzen gesammelt. Beiträge wurden sowohl als Audiodatei als auch digital über eine Word-Cloud-Software aufgezeichnet. Die Kompetenzen wurden in Kategorien kodiert und in zwei aufeinanderfolgenden Delphi-Runden rangiert, wobei die erste online und die zweite auf dem Kongress WONCA Europe (London, Juni 2022) stattfand.
Zur Untersuchung des Zusammenhangs von formativem und summativem Assessment wurden Nutzerdaten auf der digitalen medizinischen Lernplattform AMBOSS® von 195 Studierenden der Universität des Saarlandes quantitativ analysiert. Leistung in den kursbezogenen

Multiple-Choice-Fragesitzungen sowie Nutzung der integrierten Online-Lerninhalte werden miteinander und mit der Klausurnote im Fach Allgemeinmedizin korreliert. Als Referenzkohorte dienen 10.534 Studierende von 35 anderen deutschen Universitäten.

Ergebnisse

In der Delphi-Studie nahmen 30 Personen am Town Hall Meeting und 29 bzw. 31 Personen am zweistufigen Delphi-Verfahren teil. Insgesamt wurden 19 Kompetenzen der Kategorien psychomotorisch, kognitiv und affektiv durch die Autoren und Autorinnen aus den Beiträgen kategorisiert. Nach zwei Delphi-Runden wurde eine einvernehmliche Rangfolge der Kompetenzen für jede Kategorie erstellt. Die drei wichtigsten Kompetenzen für jede Kategorie blieben in beiden Runden gleich und können daher als Kernkompetenzen betrachtet werden, die angehende Allgemeinmediziner und Allgemeinmedizinerinnen in Europa als relevant für ihre Ausbildung ansehen.

Im Rahmen der quantitativen Studie zeigte sich, dass die Anzahl der für den Kurs gelesenen Lernkarten moderat mit der Leistung im formativen Assessment korrelierte ($\rho=0{,}331$, $p=0{,}005$ und $\rho=0{,}217$, $p=0{,}034$). Leistung im formativen Assessment und Klausurergebnisse korrelierten stark in der Sommersemesterkohorte ($\rho=0{,}505$, $p<0{,}001$) und moderat in der Wintersemesterkohorte ($\rho=0{,}381$, $p<0{,}001$).

Schlussfolgerung

Die Ergebnisse zeigen, dass Prinzipien der kompetenzbasierten medizinischen Ausbildung in der allgemeinmedizinischen Aus- und Weiterbildung eingesetzt werden sollten, um hochwertige Lehre sicherzustellen. Hierzu sollten Lernende, z.B. Ärzte und Ärztinnen in Weiterbildung, regelmäßig in den Entwicklungsprozess von Lernzielen in der Weiterbildung eingebunden werden. Nur so kann ein neues, kompetenzorientiertes Curriculum sichergestellt werden, das nachfolgende Generationen fundiert auf die sich dauernd wandelnde ärztliche Tätigkeit in der Primärversorgung vorbereitet. Daneben wurde gezeigt, dass die zielgerichtete Integration von Online-Lernaktivitäten in ein blended-learning Curriculum deren Nutzung verändert. Lernaktivitäten, die formatives Assessment beinhalten, können als Prüfungsvorbereitung dienen und Lehrkräften dabei helfen, Daten für Leistungsportfolios zu sammeln. Auf diesem Wege könnten die mittels „learning analytics“ gewonnen Daten in der Zukunft den Leistungsnachweis durch summatives Assessment ergänzen oder sogar ersetzen. Sie könnten zudem ein wichtiges Werkzeug sein, um Lernende mit Schwierigkeiten bereits vor dem Prüfungstermin zu identifizieren und zielgerichtet zu unterstützen.

1.2 English abstract

Introduction

In the coming years, numerous new general practitioners will have to be trained in Germany and Europe to prevent an impending shortage of general practitioners. The principle of competency-based medical education is currently highly relevant in medical education research and can help making general practice training more captivating with a higher quality. Concepts of competency-based teaching include the design of concrete learning objectives, regular feedback to learners and greater involvement of learners in curriculum planning. Within the framework of two studies, this dissertation examines the implementation of competency-based teaching in general practice. For this purpose, a Europe-wide, structured prioritization of competencies was conducted to determine which competencies are considered most important by general practice trainees. In a separate quantitative study, the relationship between performance in formative and summative assessment within a blended learning curriculum in medical studies was investigated.

Method

A structured qualitative study using the modified Delphi method was conducted to investigate the priorities for competencies in general practice training. First, suggestions for important competencies were collected in an international Town Hall meeting with young GPs at the Vasco da Gama Movement (now: European Young Family Doctors Movement) Forum (Edinburgh, January 2022). Contributions were recorded both as audio files and digitally via a word cloud software. The competencies were coded into categories and ranked in two successive Delphi rounds, the first online and the second at the conference WONCA Europe (London, June 2022).

To investigate the relationship between formative and summative assessment, user data on the AMBOSS® digital medical learning platform from 195 students at Saarland University were analyzed quantitatively. Performance in the course-related multiple-choice question sessions and use of the integrated online learning content were correlated with each other and with the exam grade in general medicine. 10534 students from 35 other German universities served as a reference cohort.

Results

In the Delphi study, there were 30 participants in the Town Hall meeting and 29 or 31 participants in the following stages of the Delphi procedure. A total of 19 competencies in the do-

mains psychomotor, cognitive and affective were categorized by the authors from the contributions. After two Delphi rounds, a consensual ranking of competencies for each domain was established. The three most important competencies for each domain remained the same in both rounds and can therefore be considered as core competencies that prospective general practitioners in Europe consider relevant for their training.

In the quantitative study, the number of learning cards read for the course was found to correlate moderately with performance in the formative assessment ($\rho=0.331$, $p=0.005$ and $\rho=0.217$, $p=0.034$). Formative assessment performance and exam scores correlated strongly in the summer semester cohort ($\rho=0.505$, $p<0.001$) and moderately in the winter semester cohort ($\rho=0.381$, $p<0.001$).

Conclusion

Results show that principles of competency-based medical education should be applied in general practice under- and postgraduate training to ensure high-quality teaching.

For this purpose, learners, e.g., general practice trainees, should be regularly involved in the development process of learning objectives for training. This way a new, competency-oriented curriculum can be ensured, that prepares future generations of doctors well for the constantly changing medical work in general practice.

In addition, it was shown that the purposeful integration of online learning activities into a general medicine blended learning curriculum changes their use by students. Learning activities that include formative assessment can serve as exam preparation and may help teachers collect data for performance portfolios. This way, data generated through ‘learning analytics’ could complement or even replace summative assessment in the future. They could also be an important tool to identify and support learners with difficulties before exams take place.

2 Introduction

2.1 The principle of competency-based medical education in general practice

In recent years, throughout Europe, the need for young general practitioners (GPs) has been constantly growing. Numerous new GPs have to be trained in the coming years in order to compensate for the declining numbers of practicing GPs, especially in rural areas [1]. According to studies, 11,000 GPs will be missing in Germany by 2035 [2]. Meanwhile, the workload of GPs is steadily increasing. The COVID-19 pandemic has challenged GPs around the world, affecting both patient care and specialty training [3]. In response to these new challenges, innovative ways must be found to interest students and young physicians in general practice (GP). It must also be ensured that the essential skills necessary for general patient care are taught successfully during medical studies and postgraduate training. To achieve these goals, there has been increasing emphasis on competency-based medical education (CBME) during the last years, not only in GP [4, 5]. The principles of CBME were first described by McGaghie et. al in 1978 [6]. Firstly, a competency-based medical curriculum creates intended learning outcomes (ILO) based on real-life skills or abilities within the medical workplace ('competencies'). Secondly, a competency-based curriculum should be designed to instruct and enable *all* learners to obtain the required competencies. Lastly, learning environments and learning techniques in CBME should be constantly tested and refined by learners [6]. Ultimately, competency is a synergy of skills, knowledge, and experience, resulting in observable ability to provide the best possible care for each individual patient [4, 7, 8].

Obtaining a competency does not depend solely on the length and type of training, and different learners may acquire competencies at different paces. That is why CBME embodies an outcome-oriented approach. By aligning learning activities in a curriculum with the desired outcomes, it is possible to ensure that all learners, regardless of their personal learning pace and individual characteristics, ultimately achieve the same learning goals [8]. It has been stated that to use CBME effectively and sustainably, the design of a 'curricular blueprint' is beneficial. A curricular blueprint entails all components of a curriculum and shows their interrelations (see figure 1). It thereby facilitates the design process of a CBME curriculum and ensures that the intended educational goals are achieved.

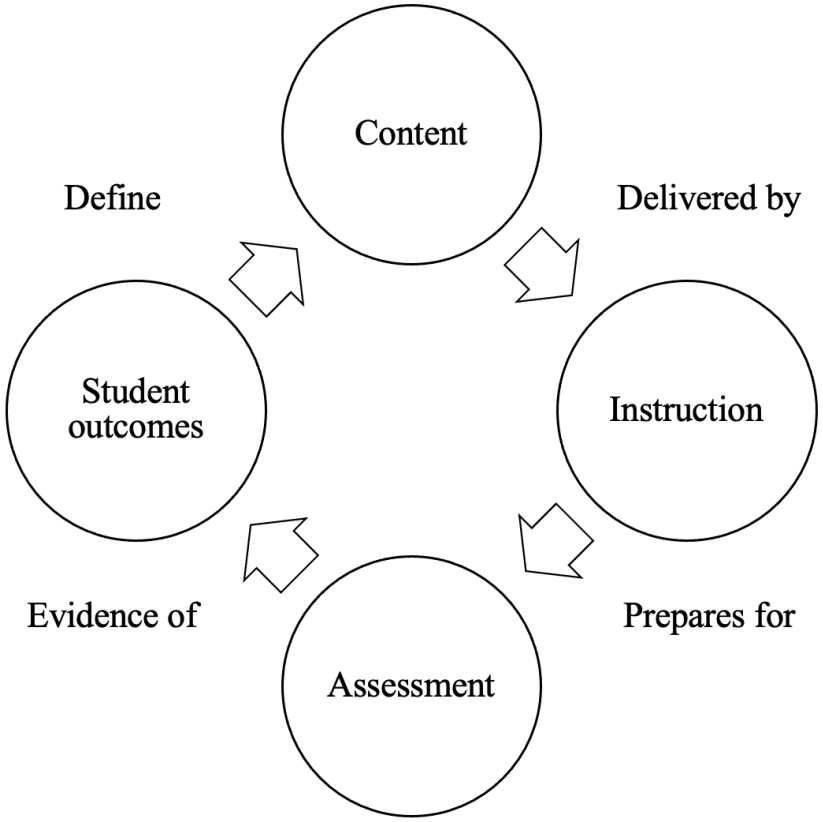

Figure 1: Scheme of a ‚curricular blueprint' [5, 9]

Consequently, one of the first steps for planning a competency-based curriculum should be the definition of ILO. A well-known example for a structured competency framework is the list of CanMEDS roles developed in 2007 [10]. Similarly, the Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Medical Specialties (ABMS) defined six core competencies necessary for the practice of medicine, namely: 1) patient care and procedure skills, 2) medical knowledge, 3) practice-based learning and improvement, 4) interpersonal and communication skills, 5) professionalism, and 6) system-based practice [11]. Frameworks like these can help to guide learners throughout the learning process and ensure the quality of learning outcomes.

2.2 Intended learning outcomes for general practice specialty training

To date, design of competency-based curriculums as well as the creation of ILO in medical education have mainly been undertaken by educators. Universities or professional organizations usually define learning objectives and determine the appropriate methods to achieve

these objectives, to then assess the learning progress. For GP, the World Organization of Doctors of Family Medicine (WONCA) and the European Academy of Teachers of General and Family Medicine (EURACT) are important stakeholders when it comes to curriculum design. Their recommendations regarding specialty training have been summarized in a 2018 paper, titled 'Educational requirements for GP specialty training'. The authors state that GP training should be outcome-based and learner-centered. Six GP core competencies for GP trainees are identified by the authors, namely (1) primary care management, (2) person-centered care, (3) specific problem-solving skills, (4) comprehensive approach, (5) community orientation, and (6) holistic modelling. These six core competencies are further specified into 12 'central characteristics of the discipline of GP', which may serve as overarching learning objectives in GP training and are depicted in the WONCA tree (figure 2).

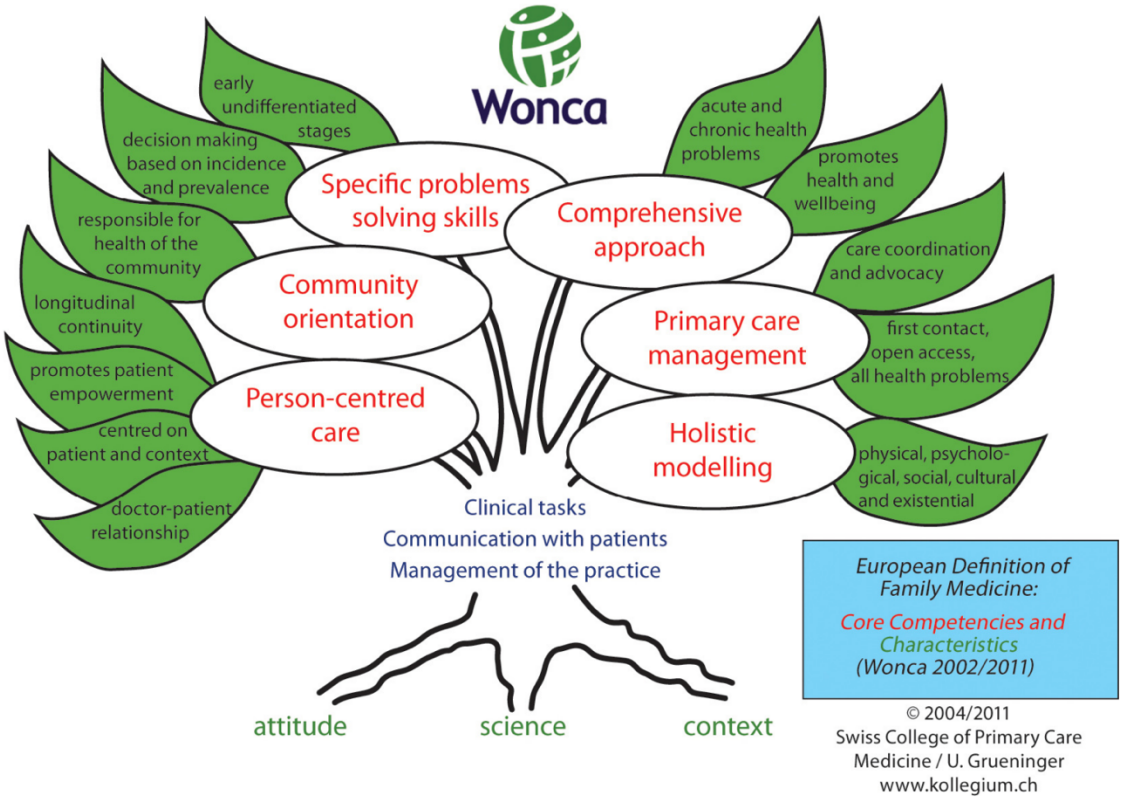

Figure 2: European Definition of Family Medicine: Core Competencies and Characteristics [12]

While there are clear statements on how educators and institutions like WONCA envision CBME in GP specialty training, there has been little integration of learners into the design process. As already stated by McGaghie et al. in 1978, structured review of learning environments and including ILO by learners themselves should be part of a competency-based curriculum [6]. For GP training, an increased inclusion of learners in the development process of

specialty training curricula has been suggested in the past but is still uncommon in practice. In a qualitative study using the modified Delphi method titled ‘What competencies do European general practice trainees value the most? A prioritisation exercise using a modified Delphi approach’ a priority list of essential competencies for future GP practice was drafted from the perspective of young GPs and GPs in training. For this purpose, young GPs at two international congresses were asked about their needs and preferences for competencies in their training. By creating a priority list structured using Bloom's Taxonomy, this Delphi study aims to answer the question of which competencies learners rate as essential to their GP profession and compare results with suggestions made by educators.

2.3 The role of assessment in competency-based medical education

As depicted in the model of a curricular blueprint, regular assessments of the learning progress are vital for CBME, since assessment drives learning. Hence, research focusing on competency-based learning should consider both ILO and assessment. Assessment is usually differentiated into two subtypes: formative and summative assessment. While summative assessment is usually undertaken at the end of a course to evaluate knowledge or proficiency, formative assessment is a part of the learning process and takes place during the course. Formative assessment is sometimes described as “assessment *for* learning” rather than “assessment *of* learning” [13]. It is an important tool for learner evaluation, serving two main purposes: identifying what has already been learned and determining areas in which the learner still has deficiencies [14]. Previous studies have shown that feedback, as used in formative assessment, can structure the learning process and have positive effects on learning itself [15, 16]. It has also been stated that formative assessment works best if implemented into a planned curriculum [17]. Formative assessment can come in different modes – e.g., through the ‘ask-tell-ask’ interview method or by verification and explanation of the correct answer after a test question [18]. In face-to-face learning activities, verbal feedback from teachers is a particularly important mode of formative assessment. In recent years, however, online teaching has been introduced in many areas of medical education. This effect has been further intensified by the necessity for social distancing due to the COVID-19 pandemic [19]. While it may appear challenging to educators to develop and apply new forms of assessment for online learning, new opportunities for assessment may arise due to the growing amount of data being collected and analyzed in online learning [20].

Even though it has been stated that the internet and innovative software may have the potential to facilitate CBME, the practice of 'learning analytics' is not yet common in medical education [21, 22]. The aim of learning analytics is to analyze and use data collected during the monitoring and evaluation of the learning process [23]. When using online learning resources, students leave a digital footprint, which can be utilized as a benchmark of learning quality and student involvement in the course [24]. Since faculty overload has been a challenge regarding assessment in CMBE, automated, digital feedback could be a major relief for educators [25]. Especially for medical education in GP, there remains a need to investigate how individual and cohort learning data can be reported back to learners and used by teachers to assess learners' academic performance.

A common tool for formative assessment in an online learning environment are multiple-choice questions (MCQs) [26, 27]. Online MCQ question banks have recently been used by many medical students not only to revise, but also to study new materials. Multiple studies have reported a universally positive attitude towards online question banks within an undergraduate and postgraduate medical student population [28, 29]. Furthermore, repeated testing with feedback appears to enhance long-term retention of information, which suggests that question banks may be an effective learning tool [30]. It has been found that including one or multiple 'tests' as part of exam preparation improves learner's exam performance. This is called the 'testing effect' [31]. For MCQ to serve as formative assessment exercises, some form of feedback has to be included, e.g., a hint and/or explanation for the correct answer [15, 17]. Being easy to conduct online and usually less expensive than other types of formative assessment, MCQs are a relevant resource for assessment in medical education.

2.4 The general practice blended-learning curriculum at Saarland University

In the winter semester of 2020/21, a new blended-learning (BL) GP curriculum was implemented for students in their fifth year of medical studies at Saarland University (UdS). Through a competency-based curriculum, students ought to acquire all necessary competencies to manage a GP patient independently. The curriculum contains three modules, each focusing on typical symptoms in GP practice: abdominal pain, febrile infection, and back pain. Emphasis is placed on students being able to identify 'red flags' that indicate an urgent need for action. Students prepare the topics with online materials, following the concept of self-

directed learning. Then, students actively train their practical competencies through patient interaction simulations.

Medical students in Germany prepare for the state examination in their fifth year. In recent years, medical online-databases have gained great popularity among students. In Germany, AMBOSS® is currently the most common online learning activity (OLA) provider in undergraduate medical education. Its database contains around 1,300 medical articles and 17,500 MCQs. As medical students usually use online learning materials they are already accustomed to, it was decided to use AMBOSS® content in complement to self-designed materials in the BL curriculum [32].

To constructively align the formative assessment tasks with the course content, specific question bank sessions have been designed from the AMBOSS® MCQ question bank. Question sessions were then integrated into course-specific online material, such as GP commentaries, podcasts, screen- and video casts and online lectures, on the curriculum's homepage. In a quantitative study titled 'The use of digital learning analytics in a blended learning family medicine curriculum at Saarland University', students' use of formative assessment exercises (AMBOSS® formative assessment sessions) and learning material (AMBOSS® learning cards) was investigated and compared with a nationwide reference cohort. This was done to answer the question whether there is a relationship between online learning behaviour, performance in formative assessment and academic success measured through summative assessment (exam grades) in a competency-based, undergraduate GP curriculum. In addition, online learning behaviour and success in formative assessment exercises of students in the BL curriculum was compared with reference cohort data.

2.5 Research question

Even though competency-based curricula are becoming more common in GP, there remains a lack of guidance on how CBME can successfully be implemented in macro-, meso- and microcurricula for different stages of under- and postgraduate GP training [33-35].
For CBME to be successful, the competencies learners should acquire must be clearly defined in a curricular blueprint. Besides, accurate assessment and effective learner feedback must be facilitated to support learners in their learning process [14]. For this reason, this dissertation will focus on a) the definition of international ILO in postgraduate GP training and b) the use and potential necessity of learning analytics in undergraduate GP training.
By analysing the use and effect of competency-based medical education in curricula in both undergraduate and postgraduate GP training, this dissertation aims to answer the following

questions: What competencies are most important for future GP practice in Europe from the perspective of young GPs and GPs in training? When online learning material is used in competency-based, undergraduate GP curricula, is there a relationship between students' learning behaviour and their success in formative and summative assessment?

3 Methods

3.1 Ethical considerations

Ethics approval for both studies described in this dissertation was obtained by Saarland medical association ethics committee on 25.09.2020 (Bu234/20) prior to study initiation. Study participation was voluntary for all participants. Consent to participation was obtained online via Google® Forms, Sli.do® or e-mail.

3.2 What competencies do European general practice trainees value the most? A prioritisation exercise using a modified Delphi approach

To answer the question which competencies GP trainees consider essential for their future work, an explorative, three-step Delphi approach was chosen. The Delphi technique is a qualitative, structured multi-stage technique. It is often used in qualitative explorative research settings and situations with little or no existing evidence [36]. By questioning a group of study participants ('experts') multiple times, the Delphi technique aims to record and transform opinions to consensus. Usually, as a first step, an open-ended questionnaire is distributed to all experts and answered anonymously to collect qualitative information. Responses are translated into general statements and summarized by researchers. Summarized responses are then fed back to participants in form of a second, quantitative questionnaire. This method is repeated until consensus among experts is obtained [37].

In this study, the open-ended questionnaire was replaced with an in-person Town Hall meeting. Town Hall meetings are informal gatherings that originally serve the purpose of citizen participation in local governance. They have recently also been a tool used within the private sector to gather employees' opinions and perspectives. The Town Hall meeting method was chosen for this study because in-person discussions have been found to increase response rates when held in early stages of the Delphi method [38]. This approach, including an in-person discussion, is also known as the modified Delphi method.

During the Town Hall meeting, the discussion was conducted face-to-face to ensure an exchange among experts. In-person discussions also enabled experts to clarify their contributions to avoid misunderstandings. Since it is recommended for Delphi studies to maintain an-

onymity during the data collection process, a digital voting tool (Sli.do®) was used to collect experts' contributions [38, 39].

Experts in this Delphi study were GPs in training or early career GPs. The definition of early career GPs was adopted from the WONCA-associated 'Vasco da Gama Movement' (VdGM; now: European Young Family Doctors Movement, EYFDM). It includes GPs during specialty training and within five years of training completion. Recruiting of experts took place at voluntary workshops during international GP conferences (VdGM Forum 2022, VdGM preconference for WONCA Europe 2022). Remaining experts were contacted because of their work as VdGM council members to ensure diversity of experts' countries of origin.

3.2.1 Town Hall meeting

As a first stage of data collection, the Town Hall meeting was held at the 7th VdGM Forum in Edinburgh on 28/01/2022 (see figure 3). The purpose of the Town Hall meeting was to draft a first list of competencies relevant for GP training. Authors structured the discussion and data collection process based on Bloom's taxonomy domains. Bloom's taxonomy, as described by Benjamin Bloom in 1956, divides competencies into the psychomotor, cognitive, and affective domains. Within these domains, Bloom further distinguishes between levels of ascending complexity [40].

Participants were informed about aim and structure of the study before the start of the Town hall meeting. The verbal discussion was digitally assisted by the word cloud/survey tool Sli.do®, which was also used to record written consent to study participation. During the meeting, participants had the opportunity to anonymously suggest competencies they considered important for GP training. Contributions by participants were simultaneously fed back to the audience via live presentation to stimulate discussion and elaboration. Follow-up questions were asked by authors to prompt further clarification of both digital and verbal contributions. The discussion was continued until theoretic saturation was reached. After the discussion, participants had the opportunity to register with their email addresses to participate in the decentralized first round of the Delphi method. They were informed they would waive anonymity by doing so.

All digital and verbal contributions were extracted into a spreadsheet (02/02/2022) and inductively coded by the authors into subgroups. The recording of the verbal discussion was reevaluated multiple times before and after coding to assure the keywords were assigned to the correct subgroup. Focused coding was done by HJ and by two other GP researchers (AP, FD) in an investigator triangulation session (10/02/2022). Striking verbal contributions

themes were transcribed verbatim. No new items were added during the analysis and experts' wording was retained wherever possible. When competencies appeared to be assigned to the wrong domain by participants, they were reassigned to the domain that authors consensually considered to be most appropriate. For the categories 'psychomotor competencies' and 'affective competencies' six subgroups were created. For the category 'cognitive competencies', due to the large number and variety of qualitative contributions, seven subgroups were created. All subgroups, containing multiple examples for each competency, were then transferred into an online survey tool (Qualtrics®, free trial version).

3.2.2 First Delphi round

The first round of the Delphi study took place as an online survey (see figure 3). The online survey was then sent out to all experts who had provided their e-mail addresses (n=13), as well as to all national representatives and executive council members of the VdGM (n=46). The survey was not distributed to individuals outside of these groups to ensure delineation of the expert panel. Participants were instructed to rank the list of GP training competencies according to their perceived importance within GP specialty training (scale 1-6 or 1-7; 1 = greatest importance, 6/7 = lowest importance). A free text form enabled participants to suggest additional competencies or modifications to the list of competencies provided. Participants were given a four-week period (11/02/2022-07/03/2022) to complete the survey. Survey responses were considered for analysis if consent was provided and if ranking was completed for at least one out of three domains of competencies. All participants received one reminder of the study via e-mail or the VdGM council WhatsApp® group.

3.2.3 Second Delphi round

For the second Delphi round, a workshop was held on 27/06/2022 at the VdGM preconference for WONCA Europe conference in London (see figure 3). All experts were reminded once (20/06/2022) via e-mail or the VdGM council WhatsApp® group to participate in the upcoming workshop. In the beginning of the workshop, a short presentation was held to inform participants about study design and preliminary study results. Results were then subjected to member checking in form of a facilitated verbal discussion. Experts' remaining open questions regarding the preliminary competency list were answered by the authors verbally during the session. Subsequently, a real time online ranking took place using Sli.do®. The scaling system used in the first Delphi round was also used in the second round (scale 1-6 or 1-7; 1 = greatest importance, 6/7 = lowest importance). Experts were asked to rank the listed

competencies according to their relevance for GP practice in their respective countries to validate the results from the first Delphi round.

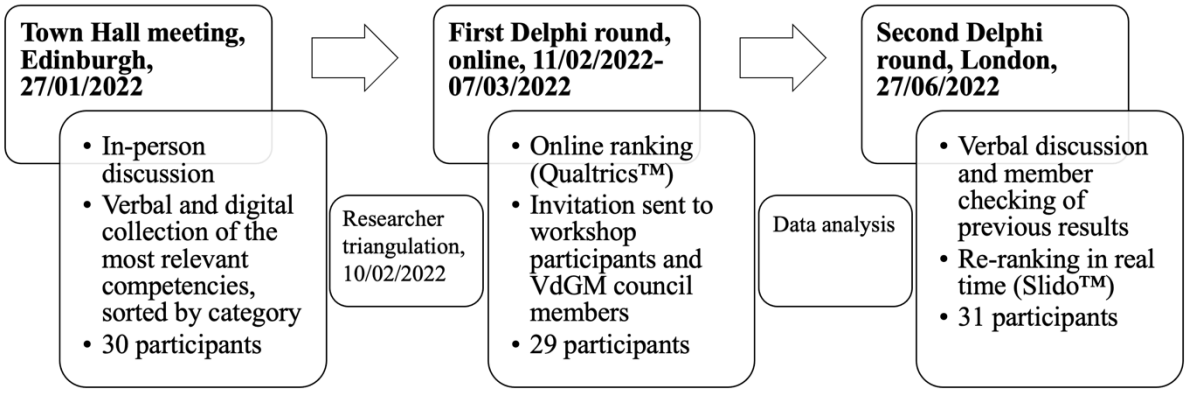

Figure 3: Scheme of the Delphi study

3.2.4 Data analysis

For quantitative data analysis, Jamovi (Version 1.6.23.0) was used. Descriptive statistics included mean and standard deviation ($\mu \pm SD$). Data was non-normally distributed (Shapiro-Wilk-Test). A degree of consensus (level of agreement) was determined as suggested in literature. A method described by De Loe et al. and adapted for a 7-point scale by Wijnen-Meijer et al. was used to calculate levels of agreement [41, 42]. A high level of agreement was assumed if 70% of the votes were cast in 2 (out of 7) adjacent scale levels or if 80% of the votes were cast in 3 (out of 7) adjacent scale levels [42]. Calculation of skewness was used to test for unidirectionality of experts' ranking.

3.3 The use of digital learning analytics in a blended learning GP curriculum at Saarland University

To investigate students' online learning behaviour in a BL GP curriculum at UdS, a quantitative approach was chosen. Participants of the study were UdS students in their fifth year who were enrolled in the compulsory GP BL curriculum during the winter semester 2020/21 or the summer semester 2021. To be included, participants had to consent to participation via an online form, have an active AMBOSS® account, provide their AMBOSS® registration email address and participate in the 60-item final exam for the GP BL curriculum.

The online part of the BL curriculum consisted partly of AMBOSS® content and partly of customized content designed and generated at the GP department of UdS. Faculty-designed

content contained multimedia learning materials such as commentaries, podcasts, screen- and videocasts and live online lectures. From the AMBOSS® database, 34 online learning cards and nine specifically designed question sessions were embedded in the BL curriculum landing page. Each question session consisted of 30 selected questions. All content was available to students during the semester through the curriculum's homepage. Students could complete the MCQ sessions at any time. They were advised to do so after studying the curriculum material. The scores of each first MCQ session attempt was considered for data analysis to minimize learning effects through repeated execution. Correct results and explanations for each question were displayed upon answering, so the MCQs could serve as formative assessment exercises.

Participants' AMBOSS® user data were recorded during the winter semester 2020/21 (01/10/20 – 28/02/21) and summer semester 2021 (01/04/21 – 31/08/21). Anonymised reference cohort user data were recorded during the winter semester 2020/21 (01/10/20 – 28/02/21). The reference cohort consisted of 10,534 students from 35 other German universities who were in their fifth year at the time of data recording. Reference cohort students were included if they had accessed at least one learning card or MCQ on AMBOSS® during the respective timeframe and if they had been provided access to AMBOSS® by their university free of charge. A power analysis to estimate the cohort size for this study was performed using G*power. Power analysis for correlation with an effect size of 0.5 and an $\alpha=0.05$ suggested an estimated sample size of 34. For an effect size of 0.3 and an $\alpha=0.05$, a total sample size of 109 was suggested.

3.3.1 Data analysis

Quantitative data analysis was performed using Jamovi (Version 1.6.23.0). Normality was tested using Shapiro-Wilk-Test. All analysed data was non-normally distributed. Descriptive analysis included mean, median and standard deviation. To test for differences in data between summer and winter semester cohort and between UdS students and the reference cohort, Mann-Whitney U-Test (U) was used. Correlation analysis was conducted using Spearman's rho (ρ). Effect sizes of $\rho<0.3$ were considered a small effect; $\rho=0.3$-0.5 a moderate effect and $\rho>0.5$ a large effect.

4 Results

4.1 What competencies do European general practice trainees value the most? A prioritisation exercise using a modified Delphi approach

4.1.1 Sociodemographic data

30 individuals attended the first Town Hall discussion on 28/01/2022. 23 (77%) actively participated in the digital data collection process. All 23 participants reported working in GP. At the time of the event, six participants (26%) were in specialty training, eleven participants (48%) within the first five years after specialty training and six participants (26%) more than five years after specialty training. No university students or doctors in their preregistration/foundation year took part in the survey.

In the first Delphi round, 29 experts completed the survey and were considered for data analysis. Panel members were from 22 different countries (see figure 4). The most common countries of origin were France (n=3, 10.3%), UK (n=3, 10.3%) and Spain (n=3, 10.3%). Most participants were within five years after GP specialty training (n=15, 54%). About a third of participants were currently in GP specialty training (n=10, 36%), while 11% (n=3) had finished GP specialty training more than 5 years ago. No university students or doctors in their preregistration/foundation year took part in the voting. One participant did not provide information on their level of training.

36 individuals were physically present at the workshop for the second Delphi round on 30/06/2022. 33 (92%) actively participated in the digital data collection process. Panel members were from 17 different countries (see figure 4). The most common countries of origin were the Netherlands (n=6, 18%) and the UK (n=5, 15%). Most participants were in GP specialty training (n=19, 58%). About a third of participants were 5 years within GP specialty training (n=12, 36%), while 6% (n=2) had finished GP specialty training more than 5 years ago. No university students or doctors in their preregistration/foundation year took part in the voting.

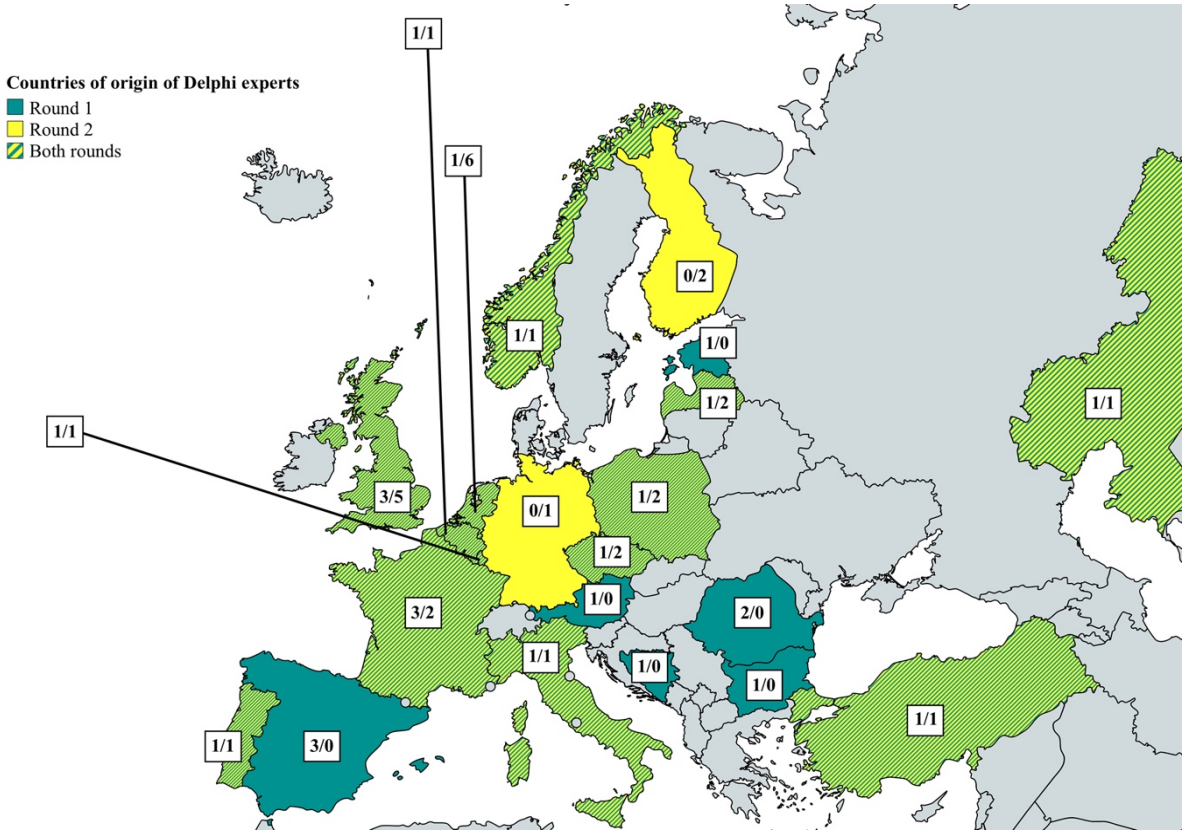

Figure 4: Countries of origin for participants in the Delphi rounds and number of participants per round (Round 1/Round 2). Not on map: Kyrgyzstan (1/0), Malta (1/0), Israel (1/1). Created using www.mapchart.net [43].

4.1.2 Town Hall meeting

The participants suggested a total of 139 competencies via Sli.do® during the Town Hall meeting. These were coded into 19 subgroups by the authors, duplications were removed. For subgroups and examples for the different subgroups, see tables 1-3.

Subgroup	Examples
General physical examination	general physical examination
Specific examination skills	ear/nose/throat-examination, fundoscopy, dermatological-, musculoskeletal-, or gynecological examination
Documentation & digital skills	typing, documenting succinctly, use of technology
Using and analysing diagnostic tools	spirometry, ECG, blood samples
Interventional skills	wound care, minor surgery, fitting contraception, joint injection
Conducting imaging/POC diagnostics	ultrasound

Table 1: Subgroups and examples for the domain of 'psychomotor competencies'

Subgroup	Examples
Individualised care	translating guidelines to individual care, supporting patient self-management, health coaching skills, integrated care
Medication & prescribing	prescribing and deprescribing, polypharmacy, pharmaceutical interactions
Condition specific knowledge	emergency medicine, obstetrics, pediatrics, surgery, gynecology
Mental health skills	mental health assessment, learning disabilities assessment, basic psychotherapy, psychiatric status
Health systems & finance	knowledge about healthcare systems, digital skills, payroll, clinical coding, time management, referral criteria, triage
Infectious diseases	epidemiology of emerging diseases, vaccines
Imaging interpretation	interpreting x-ray results

Table 2: Subgroups and examples for the domain of 'cognitive competencies'

Subgroup	Examples
Communication skills	motivational interviewing, active listening, breaking bad news, negotiation, assertiveness, empathy, coaching
Establishing doctor/patient relationship	empowering patients, patient focused care, patient perspective
Managing wellbeing	saying 'no', self-care, recognizing limitations, emotional management, advocacy for the GP profession
Sensitivity towards differences	creating a safe space, 'non-judgementalism', open mindset, awareness of diversity, awareness of own prejudices, equality and diversity
Teamwork & time management	discussion of errors, teaching, teamwork, time management, leadership
Clinical quality improvement	dealing with 'difficult' cases, managing patients' expectations, managing uncertainty

Table 3: Subgroups and examples for the domain of 'affective competencies'

4.1.3 First Delphi round

Psychomotor competencies

28 participants took part in the ranking of psychomotor competencies. 'General physical examination skills' had the lowest average competency score (1.32 ± 0.945) and the greatest skewness (3.25). This indicates experts consider the physical examination to be highly relevant for GP training. 'Specific examination skills' was ranked in second place, and 'documentation and digital skills' in third place (see table 4). Overall, skewness was positive for competencies ranked with high importance and low for competencies ranked with low importance, indicating unidirectionality. A high level of agreement was found for the categories 'general physical examination' and 'conducting imaging/POC diagnostics'. A medium level of agreement was found for all other categories.

Cognitive competencies

29 participants took part in the ranking of cognitive competencies. 'Individualised care' was ranked as the most important cognitive competency with a medium rank of 1.86 ± 1.43 (see table 4). It was followed by 'medication and prescribing' and 'clinical/condition specific knowledge'. All other four categories of cognitive competencies were assigned a lower average rank. A high level of agreement was found for three categories and a medium level of agreement was found for one category. No agreement could be found for the categories health system and finance, condition specific knowledge and mental health skills. Accordingly, these categories showed a skewness close to zero.

Affective competencies

29 participants took part in the ranking of affective competencies. 'Communication skills' were ranked as the most important cognitive competency with a medium rank of 1.76 ± 0.95, followed by 'doctor/patient relationship'.

Despite the great relevance of the topic 'Managing personal wellbeing' during the discussion at the Town Hall meeting in Edinburgh, this competency was only assigned an average rank of 3.62 (see table 4). A high level of agreement was found for 50% (n=3) of the affective competencies, a medium level of agreement for the other 50% (n=3).

Competency	Mean	Median	SD	LoA	Skewness
PC: General physical examination	1.32	1	0.945	High	3.25
PC: Specific examination skills	2.96	2	1.6	Medium	0.942
PC: Diagnostic tools	3.39	3	1.26	Medium	0.622
PC: Interventional skills	3.79	4	1.26	Medium	0.436
PC: Documentation & digital skills	4.61	5	1.29	Medium	-0.537
PC: Conducting imaging/POC diagn.	4.93	5	1.09	High	-1.72
CC: Individualised care	1.86	1	1.43	High	2.22
CC: Medication & prescribing	2.69	3	1.11	High	0.844
CC: Condition specific knowledge	3.97	4	1.86	None	-0.268
CC: Applying mental health skills	4.31	4	1.83	None	0.251
CC: Health systems & finance	4.34	4	1.95	None	-0.0059
CC: Interpretation of imaging	5.28	5	1.22	Medium	-0.32
CC: Infectious diseases	5.55	6	1.64	High	-1.62
AC: Communication skills	1.76	1	0.951	High	1.06
AC: Establishing doctor/patient relation-ship	2.1	2	0.86	High	0.154
AC: Managing one's own wellbeing	3.62	4	1.37	Medium	-0.315
AC: Clinical quality improvement	4.14	4	1.41	Medium	-0.675
AC: Teamwork & time management	4.28	5	1.39	Medium	-0.708
AC: Sensitivity towards differences	5.1	6	1.35	High	-1.43

Table 4: Psychomotor competencies (PC), cognitive competencies (CC), affective competencies (AC) and their rank, Level of Agreement (LoA) and skewness in Round 1 of the Delphi study. Green color: High or medium LoA. Yellow color: low LoA. Red color: no consensus [43].

4.1.4 Second Delphi round

Psychomotor competencies

31 participants took part in the ranking of psychomotor competencies. The two competencies considered most important for GP training by experts in round one were confirmed in the second round (see table 5). Unlike the first Delphi round, ‘documentation and digital skills’ was not considered one of the top three competencies and was moved to the fifth rank. Except for this change, the order of the competencies stayed the same. Again, skewness was positive for competencies ranked with high importance and low for competencies ranked with low importance respectively, indicating unidirectionality. Identically to the first round, a high level of agreement was found for the categories ‘general physical examination’ and ‘conducting imaging/POC diagnostics’. A medium level of agreement was found for all other categories.

Cognitive competencies

31 participants took part in the ranking of cognitive competencies. Results from the first Delphi round were mostly confirmed: the first five ranks of competencies in this domain stayed the same. There was a change in the order regarding the lower scored items. Unlike in the first Delphi round, the category ‘infectious diseases’ was now considered the least important cognitive competency out of the seven competencies ranked. Only low levels of agreement could be found for the categories ‘health system and finance’ and ‘condition specific knowledge’. Accordingly, these categories showed a skewness close to zero. A high or medium level of agreement was calculated for all other categories.

Affective competencies

31 participants took part in the ranking of affective competencies. Ranks of all six competencies from the first Delphi round were confirmed: ‘Communication skills’ was ranked as the most important affective competency, followed by ‘doctor/patient relationship’ and ‘managing personal wellbeing’. ‘Clinical quality improvement’ was considered slightly more important by experts in round two and moved up to rank four. The remaining two competencies had the same mean (see table 5), hence shared the fourth rank. Most categories showed low levels of agreement. Exceptions were ‘teamwork and time management’ (medium LoA) and ‘managing one’s own wellbeing’ (no agreement).

Competency	Mean	Median	SD	LoA	Skewness
PC: General physical examination	1.43	1	1.07	High	2.79
PC: Specific examination skills	3.00	3	1.26	Medium	0.777
PC: Diagnostic tools	3.30	3	0.952	High	0.878
PC: Interventional skills	3.87	4	1.61	Medium	-0.667
PC: Documentation & digital skills	4.10	4	1.49	Low	-0.182
PC: Conducting imaging/POC diagn.	5.30	6	1.02	High	-1.7
CC: Individualised care	2.27	1.5	1.62	High	2.22
CC: Medication & prescribing	2.33	2	1.12	High	0.844
CC: Condition specific knowledge	2.97	3	1.69	Low	-0.268
CC: Applying mental health skills	4.17	4	1.26	Medium	0.106
CC: Health systems & finance	5.07	6	1.74	Low	-0.698
CC: Interpretation of imaging	5.53	6	1.66	Medium	-0.32
CC: Infectious diseases	5.67	6	1.12	High	-1.62
AC: Communication skills	1.93	2	1.16	Low	1.17
AC: Establishing doctor/patient relation-ship	2.97	3	1.43	Low	0.223
AC: Managing one's own wellbeing	3.59	4	1.64	None	0.0455
AC: Clinical quality improvement	4.07	5	1.85	Low	-0.617
AC: Teamwork & time management	4.17	4	1.31	Medium	-0.239
AC: Sensitivity towards differences	4.28	4	1.62	Low	-0.427

Table 5: Psychomotor competencies (PC), cognitive competencies (CC), affective competencies (AC) and their rank, Level of Agreement (LoA) and skewness in Round 2 of the Delphi study. Green color: High or medium LoA. Yellow color: low LoA. Red color: no consensus [43].

4.2 The use of digital learning analytics in a blended learning GP curriculum at Saarland University

4.2.1 Sociodemographic data

86 students from winter semester 2020/21 and 109 students from the summer semester 2021 took part in the study. Mean age of participants in the winter semester was 26.3 ± 4.61 years, while mean age of participants in the summer semester was 24.2 years ± 2.35 years. The gender distribution was similar in both semesters. Among participants in the winter semester, 58.2% (n=32) of the participants were female and 41.8% (n=23) were male. In the summer semester cohort, 58.3% (n=63) of the participants were female and 41.7% (n=45) were male.

4.2.2 Digital learning behaviour

Digital learning behaviour data was available for 90% (n=98) of summer semester students and 85% (n=73) of winter semester students. Out of 34 AMBOSS® learning cards included in the BL curriculum, students accessed on average 31 ± 4.7 in the summer semester and 30 ± 5.0 in the winter semester (see table 6 and 7). The difference between these numbers was not significant (U=3046, p=0.092). Regarding the numbers of accesses for the 34 selected learning cards, it was found that summer semester students accessed the cards significantly more often than winter semester students (185 ± 143 vs. 131 ± 71.7; U=2726, p=0.008). For UdS students in the BL curriculum, a strong correlation was found between the number of selected learning cards read and the frequency of accesses for the selected learning cards, both in the summer semester (ρ=0.525, p<0.001, see appendix: table 10) and the winter semester (ρ=0.632, p<0.001, see appendix: table 11).

Students from the reference cohort that used AMBOSS® but did not take part in the GP BL curriculum read significantly fewer of the selected learning cards, namely 10 out of 34 cards (SD: 8.28; U= 48200, p<0.001). On average, the reference cohort also accessed the selected learning cards significantly less frequently than the UdS cohort (38 ± 54.2 vs. 162 ± 120; U=157892; p=0.001).

Looking at the activity across all learning cards available on AMBOSS®, it can be stated that UdS students in the summer semester have accessed significantly more learning cards than students in the winter semester (665 ± 348 vs. 313 ± 240; U=1561; p=0.001).

	Selected learning cards read	Access time selected cards	Total number of questions answered	Total number of learning cards read
N	98	98	98	98
Missing	11	11	11	11
Mean	31	185	6496	665
Median	32	144	7127	887
SD	4.66	143	4574	348
Minimum	6	13	84	37
Maximum	34	933	24014	1082

Table 6: Descriptive analysis of digital learning behaviour for UdS students in the summer semester 2021 [44].

	Selected learning cards read	Access time selected cards	Total number of questions answered	Total number of learning cards read
N	73	73	74	73
Missing	13	13	12	13
Mean	30	131	1276	313
Median	31	116	335	213
SD	5.02	71.7	1977	240
Minimum	10	21	0	32
Maximum	34	290	6853	863

Table 7: Descriptive analysis of digital learning behaviour for UdS students in the winter semester 2020/21 [44].

4.2.3 Formative assessment performance

Regarding formative assessment, two different variables were considered: the questions answered for the BL curriculum and the total number of questions answered on AMBOSS®.

AMBOSS® MCQ selected for the BL curriculum

80 (93,02%) students from the winter semester and 106 (97,25%) from the summer semester completed at least one question from the MCQ sessions designed for the BL curriculum. Out of 180 multiple choice questions provided, summer semester students answered on average

166 (92%) multiple choice questions and winter semester students 172 (96%) questions. The mean relative result score for the initial session was called the 'formative assessment score'. The formative assessment score was 0.860 ± 0.0841 for summer semester students and 0.789 ± 0.112 for winter semester students, with 1.0 being the maximum score achievable. The lowest formative assessment score achieved was 0.428 in the summer semester and 0.365 in the winter semester. The highest formative assessment score achieved was 1.0 in the summer semester and 0.994 in the winter semester. The difference between average formative assessment scores of the summer- and winter semester cohort was significant (U=2452, p=<0.001).

All MCQ on AMBOSS®

In total, summer semester students answered on average 6496 questions in total with a success rate of 81%, while winter semester students only answered on average 1276 questions with a success rate of 79%. This means summer semester students answered significantly more questions on AMBOSS® than winter semester students (6496 ± 4574 vs. 1276 ± 1977; U=856, p<0,001).

In both UdS semester cohorts, the overall question success rate on AMBOSS® correlated positively with formative assessment scores. This correlation was strong for the summer semester cohort (ρ = 0.789, p<0.001, see appendix: table 10) and moderate for the winter semester cohort (ρ = 0.419, p<0.001, see appendix: table 11). There was a strong correlation (ρ = 0,568, p<0.001, table 8) between the overall question success rate and the exam scores of students in the summer semester. This association could not be found for the winter semester cohort.

4.2.4 Formative assessment performance and academic success

99% (n=85) of winter semester- and 98% (n=107) of summer semester students took part in the final exam of the GP BL curriculum. The maximum score was 60 points. The mean score was 51.5 ± 6.22 in the winter semester and 51.7 ± 4.64 in the summer semester. There was a correlation between formative assessment scores and exam results in both cohorts. This correlation was strong (ρ= 0.505, p<0.001, table 8) in the summer semester cohort, and moderate (ρ = 0.381, p<0.001, table 9) in the winter semester cohort. There was also a moderate correlation between scores in the first state exam grade and formative assessment scores. This correlation was weaker in the summer semester cohort (ρ=0.332, p<0.01, table 8) than in the winter semester cohort (ρ=0.414, p<0.01, table 9).

		Exam score	First state exam grade	Question success rate	Formative assessment score
Exam score	Spearman's rho	—		0.568	0.505
	p-value	—		<.001	<.001
First state exam grade	Spearman's rho	-0.439	—	-0.420	-0.332
	p-value	< .001	—	<.001	<.001
Question success rate	Spearman's rho			—	
	p-value			—	
Formative assessment score	Spearman's rho			0.789	—
	p-value			<.001	—

Table 8: Correlation matrix: online learning behaviour, formative assessment scores and academic performance for UdS students in the summer semester 2021 [44].

		Exam score	First state exam grade	Question success rate	Formative assessment score
Exam score	Spearman's rho	—		0.244	0.381
	p-value	—		0.060	<.001
First state exam grade	Spearman's rho	-0.308	—	-0.470	-0.414
	p-value	0.005	—	<.001	<.001
Question success rate	Spearman's rho			—	0.789
	p-value			—	<0.001
Formative assessment score	Spearman's rho			0.419	—
	p-value			0.001	—

Table 9: Correlation matrix: online learning behaviour, formative assessment scores and academic performance for UdS students in the winter semester 2020/21 [44].

5 Discussion

5.1 The use of digital learning analytics in a blended learning GP curriculum at Saarland University

In this study, the method of learning analytics was used to investigate the relationship between students' use of online learning resources, formative assessment exercises and their academic success measured in GP exam scores. Results show that online learning content is used more extensively when embedded in a BL curriculum. Greater use of learning content seems to be associated with better scores in formative assessment exercises. High scores in formative assessment correlate with high scores in summative assessment (GP curriculum final exam).

The analysis of the digital learning behaviour of the fifth-year students at UdS reveals a clear difference between the summer and winter semester cohort. There is a greater use of OLA and a stronger performance in formative assessment exercises for summer semester students. This may be explained by the fact that students at UdS usually prepare for the state examinations during the summer if they stay within the standard study schedule. Since many students in Germany use AMBOSS® learning cards and MCQs for state exam preparation, summer semester students may have been already familiar with these forms of online learning. This is a potential explanation for the fact that a strong correlation between exam scores and the overall question success rate on AMBOSS® could only be found for the summer semester cohort, but not for the winter semester cohort.

When analysing the use of learning cards selected for the GP BL curriculum, it becomes apparent that UdS students read a higher amount of selected learning cards than students in the reference cohort. They also accessed the selected learning cards more often. These results indicate that students' online learning efforts can be enhanced if faculty pre-selects OLA and then specifically incorporates them into a BL curriculum. Since there was a strong correlation between the total number of learning cards read and the frequency of accesses for those learning cards, it may be assumed that students who read a lot of learning cards also studied them with a higher frequency.

An interesting result of this study, which confirms evidence from previous literature, is the finding that formative assessment scores correlate strongly with FM subject exam scores and moderately with scores for the first state exam. This suggests that formative assessment could

assist faculty in identifying weak students already during the course. This way, weak students could receive extra support and possibly be saved from ‘falling through the cracks’. In the future, these findings could also help place a stronger emphasis on formative assessment in German medical education. This way, it may be possible to progress from summative assessment as the sole marker of students’ performance to various sources of data as a performance record.

Overall, the results suggest that integrating OLA into a curriculum encourages its use by students. The use and acceptance of OLA, including online formative assessment, appear to be higher before high stake exams, like state exams. It also seems that formative assessment is a good indicator of academic performance, especially when used intensively by students.

5.2 What competencies do European general practice trainees value the most? A prioritisation exercise using a modified Delphi approach

In this study, for the first time, European GP trainees have been questioned about their preferences regarding competencies in GP training. By applying the Delphi method and conducting an open data collection at the beginning of the study, participants were free to submit any competencies and were not restricted by suggestions from faculty or educators. This way, rather than replicating or redesigning existing GP training curricula, the study aimed to assess young GPs opinions. The drafted list could enable GP educators and curriculum planners to better adapt training to the needs of GP trainees.

Evidently, competencies are often multi-layered, since they require the combination of knowledge, skills and attitude in a way that enables to deal with a situation or patient [4, 7, 8]. For this reason, competencies can rarely be clearly assigned to one of the domains used in this study. The view of workplace activities through the lens of Bloom's Taxonomy remains an artificial one. Bearing this circumstance in mind, it was decided to use the concept of Bloom’s Taxonomy in this study to enable learners without much theoretical knowledge of medical education to gain better understanding of the topic.

Competencies cannot be taught well through unstructured, isolated learning activities alone. A key element of planning a curriculum is to allocate competencies to ILO and constructively align them in a curricular blueprint [5, 7]. This circumstance should be apprehended when considering the results presented in this study.

For the domain of psychomotor competencies, it is noticeable that, unlike for the other two domains (*affective* and *cognitive*), consensus was reached for almost all categories in both

rounds. From the perspective of young GPs, ‘general physical examination’ is clearly considered the most relevant competency in GP training, followed by other examination techniques. Interventional und diagnostic skills also seem to be considered important by Delphi experts in this study. This may suggest that in the opinion of young GPs, becoming a ‘hands-on examiner’ is an important outcome of GP training. When implementing the training of ‘hands-on skills’ into a curriculum, it may be necessary to consider local factors. In a rural work environment, skills like ‘condition specific knowledge’ or ‘imaging interpretation skills’ might be of higher importance for GPs [45, 46].

In this study, ‘individualised care’ was ranked on first place in the cognitive domain. It can also be considered as the most complex competency listed according to the principle of Bloom’s taxonomy [40]. Besides the activation of factual knowledge, it requires a critical evaluation of the guidelines, assessment of patient preferences and needs as well as analysis of treatment options. The apparent great need of young GPs to acquire such a complex competency suggests that they may be more interested in the application of knowledge than in the aggregation of detailed theoretical knowledge itself.

It is striking that in both Delphi rounds the cognitive competency regarding ‘health systems & finance’ was assigned a medium rank. Apparently, this topic has not been sufficiently covered in GP trainings in Europe so far. In the context of increasing bureaucratisation in GP practice, it seems to be important for young GPs to see themselves not only as medical experts, but also as mediators between patients and the health system. The low level of agreement suggests that there is no consensus on the exact relevance of this topic among European GPs yet.

Regarding affective competencies, there appears to be a clear consensus that communication skills and establishing a doctor-patient relationship are among the most important affective competencies in GP training. When evaluating the in-person discussions in Edinburgh and London, it becomes apparent that the affective domain was the most controversial of all three domains. This is also reflected in low levels of agreement in the second Delphi round. The underlying reason could be that personal characteristics, cultural background and experiences of individual GPs, as well as the ‘hidden curriculum’, may influence experts’ attitudes towards affective competencies [47].

Of course, this study can only be a snapshot of the needs of today’s European GP trainees. Their preferences towards GP training may change over time, and regular reassessment of findings would be necessary to ensure their accuracy [48].

5.3 Results in the context of literature

Research in this dissertation mainly focuses on two parts of the curricular blueprint as described in literature: 1) the design of ILO and 2) measuring learning progress.

The process of ILO design has been described and investigated in previous studies. Some of these also use the Delphi method but target other interest groups like nurses or undergraduate medical students [49, 50]. Various studies address ILO design in GP specialty training, but for specific national contexts or from the perspective of other stakeholders than GP trainees (e.g., educators, institutions). In 2021, Song et al. have asked Korean GPs to rate topics and procedures by their importance, to design a list of essential facets for residency training in Korean family medicine. A list of 153 topics and 81 procedures was drafted [51]. De Vocht et al. (2022) used semi-structured interviews to question 22 Belgian first year GPs regarding their preferred learning outcomes for during their hospital rotation. The study confirmed the findings of our Delphi survey that local factors influence the preferences of GP trainees regarding their specialty training [52]. A 2021 qualitative study from the Netherlands used focus group interviews with Dutch GP trainees to investigate their perceptions on affective skills. Like participants in the Delphi study, Dutch GP trainees requested affective skills like clinical empathy to be included in the training curriculum [53]. Interestingly, in our Delphi study, the domain of affective competencies showed the lowest rate of consensus overall. This could be caused by the circumstance that out of all competency domains, the affective domain has received the least attention in past research and curriculum design [54]. Many European curricula for general practitioners do not include explicit training in affective competencies yet. However, the lively discussions in the Delphi study have shown that the GP trainees' interest in this topic is very high. Future studies should investigate their preferences further, especially on a pan-European level.

Since the Delphi study among GP trainees focuses on continuing education in Europe, it is important to refer to previous publications of the responsible curriculum planners. To investigate educational requirements, EURACT has already conducted a needs-analysis among both novice and expert European GP educators [55]. The authors state that even coming from different cultural backgrounds and working in different healthcare systems, participants reported similar problems and needs regarding their role as GP educators. By including European GPs in training into the curriculum design discussion, our study provides the complement to these findings from the perspective of another stakeholder. In 2017, EURACT council members have also stated that establishing common ILO for specialty training is necessary to strength-

en general practice within European countries [56]. By making GP trainees prioritize and help improve ILO on a pan-European level, this Delphi study aims to fill this gap.

Interestingly, there is a lot of overlap of Delphi study results with existing WONCA and EURACT recommendations regarding GP specialty training. As depicted in the WONCA tree, these stakeholders emphasize the qualities of 'communication skills', 'doctor-patient relationship' and 'individualised care' under 'person-centered care', all of which were also highlighted by GP trainees in our study. WONCA and EURACT further name the competency 'holistic modelling' as essential. It is being defined as dealing with 'problems in their physical, psychological, social, cultural and existential dimensions' [12]. This is consistent with the competency category 'sensitivity towards differences' in our Delphi study. This comparison shows that all stakeholders involved in GP curriculum design ultimately have similar interests. Including learners in the design of curricula and ILO is therefore not only necessary but also realistically feasible.

Assessment in CBME has been a highly relevant topic in the medical education research in recent years. Especially since the shift to online learning has been accelerated by the global COVID-19 pandemic, medical students readily and frequently use OLA to study and assess their own learning progress [57, 58]. Self-assessment tools are particularly popular in the context of self-directed learning, as implemented in the UdS GP BL curriculum [59]. Consistent with our findings, previous studies have found a relationship between students' performance in formative assessment exercises and their summative assessment scores. This correlation has been demonstrated, among others, for undergraduate medical programmes in Saudi Arabia, USA and Australia, and for different clinical subjects [60-62]. Our study confirms that this interrelation also applies to learners in an undergraduate GP curriculum in Germany.

In previous research, most OLA and online formative assessment exercises were developed uniquely for university-specific BL curricula [63-66]. The choice to include AMBOSS® content in the UdS BL curriculum is based on the intention to reduce the learning burden for students. The presentation form of learning material, as well as the adaption to it, can cause 'extraneous cognitive load' according to the Cognitive Load Theory [67]. A study at UdS has found that students' motivation can by increased by keeping the extraneous load as low as possible [68]. Literature also indicates that medical students prefer to use study material they are already familiar with [32]. Since most year 5 students in Germany are already experienced in using AMBOSS®, the familiar structure of the learning and assessment material may break down barriers and enhance the effect of online learning.

As shown by Prober et. al in 2013, students tend to be dissatisfied with course curricula that fail to reflect the expectation horizon for standardized national examinations, and it can even drive them to prefer third-party study material over university-specific learning resources [69]. Besides the fact that using an already existing online learning system can save the faculty financial and human resources, it may also increase student satisfaction when curricula are aligned with national learning objectives. This is especially relevant when, like in Germany, the national state examinations form the lowest common denominator of the different coexisting medical school curricula [69].

When making these considerations, one should not overlook the fact that undergraduate medical education consists of more than preparation for the state examination. Competencies arise from attitude, knowledge, and skill. They cannot be adequately taught without including all these three sub-aspects. So far, online learning programmes have only been able to adequately teach and assess the domain of knowledge, and mostly rely on quantitative data. Ideally, assessment systems should include both quantitative and qualitative data [70]. Despite these limitations, the effect of formative assessment should not be underestimated, as 'assessment drives learning'.

Another point of criticism in the existing literature is the one-dimensionality of MCQ as an assessment tool. There is a risk that learners answer questions correctly only through the effect of recognition. This is referred to as 'cueing' [71]. Within a current research cooperation with the German Institute for Medical Examination Questions (IMPP), methods to avoid such effects are being investigated at UdS. One tool to help to simulate clinical decision-making are key-feature MCQ sequences. Within the sequence, it is not possible to revisit a previous question once a subsequent question has been opened. This prevents additional information given in later items from influencing the decision-making process [72]. Through key-feature questions, the assessment of a learners' clinical reasoning is possible, and 'cueing' may be reduced [71, 73].

Additional research on the MCQ format is necessary so that it reflects skills from daily clinical practice even better in the future. Besides, future studies should investigate if the findings of these study are true for other forms of formative and summative assessment.

Although the term learning analytics is not yet very common in medical education, the collection and analysis of online learning data from medical students has been carried out in many previous studies [74-76]. As exemplified in this study, implementing learning analytics throughout medical schools would enable educators to monitor learners both on individual

and cohort levels. This would be particularly useful in times of mandatory remote learning, e.g., due to social distancing.

Additionally, there has long been a tendency in medical education research to shift away from summative assessment as the only proof of performance. Learning analytics could support these efforts. As stated by Lockyer et al. in 2017, digital technology may play a role in both formative and summative assessment, in low and high stakes contexts [21]. By feeding into e-portfolios, this data may provide the learner with a longitudinal view of their past learning and thereby eventually even replace summative assessment in the future [21, 77-81].

5.4 Limitations

5.4.1 What competencies do European general practice trainees value the most? A prioritisation exercise using a modified Delphi approach

Overall, a selection bias can be assumed regarding the experts in this Delphi study. Presumably, participants were mainly European GPs with a special interest in CBME. In this study, this aspect may possibly be seen as beneficial, as participants with a high interest in CBME are also more likely to make valuable contributions to the discussion.

Participants also had to have time and financial resources to travel to international conferences, which probably limited the diversity of the expert panel. Since data collection took place at conferences held in the UK, a high number of experts came from Western European countries. Participants had to be proficient English speakers, since both workshops and the online survey were only available in English. Future studies should identify the preferences of financially disadvantaged young GPs and young GPs from Eastern Europe, since these stakeholders have not been sufficiently represented in this study. This could be done by diversifying the sites of data collection or by conducting multilingual studies.

Regarding the stage of training of the Delphi experts, the subgroup of foundation year doctors (UK) or first-year physicians is underrepresented in this study. Future studies should further investigate this group's opinion, as their needs regarding GP specialty training might differ from GPs that have already been working in a practice for multiple years.

When regarding this study’s results, it is important to remember that the drafted list is not exhaustive. The aim was to create a priority list. It can only cover a certain part of competencies required in the field of GP. The ranking format may leave readers with the impression that competencies ranked in the lower positions are not important. However, the fact that they

were suggested in the pre-Delphi discussion and included in the list of results already indicates that they are of high relevance for young GPs.

Lastly, this study is limited by its exploratory approach. Although the Delphi technique has often proven successful in exploratory settings and the results should be considered embedded in the existing literature, it remains unclear to what extent the results of this study can be generalised. Further surveys among young GPs, both in European and non-European settings, should be conducted to verify the results of this study.

5.4.2 The use of digital learning analytics in a blended learning GP curriculum at Saarland University

Four students in the winter semester had not correctly submitted their AMBOSS® registration email address. Due to this circumstance, their user data could not be matched with consent forms. Without clear matching, participants were excluded from the study.

All MCQ in the AMBOSS® database and therefore all questions in the formative exercises for the BL are old state exam questions provided by the German Institute for Medical and Pharmaceutical Examination Questions (IMPP). Because of an ongoing research cooperation between the department for GP at UdS and IMPP, the GP final exam is also based on old state exam questions. Even though it was ascertained that none of the questions were used in both the formative assessment and the GP final exam, this circumstance may have influenced study results.

Although MCQ are a well-known and popular tool among educators, it is evident that they can only assess one element of the skills acquired throughout medical school. Apart from teaching theoretical knowledge, the GP BL curriculum at UdS includes on-site simulations to teach students practical competencies such as examination techniques or communication and patient management skills. These skills cannot be assessed well with MCQs. It remains unclear how the increase in competencies in the non-theoretical domains influence the GP exam grade as well as students' skilfulness overall within this GP curriculum. Other types of assessment, such as OSCE exams, may be better suited to measure these non-cognitive skills, and should be included in potential future online-portfolios, but they are often costly and difficult to administer online in times of distant learning.

Although only fifth-year students were considered for the reference cohort, the curricula of the various German medical schools differ. It remains unclear which subjects the students of the reference cohort were taught in their fifth year and whether and to what extent digital learning was offered during the study period.

Lastly, this study only addresses data from a GP curriculum. Overarching, cross-disciplinary types of online learning and formative assessment should be developed and implemented in the future, so the 'whole' of learning in medical school can be captured [17]. This may require more cross-disciplinary collaboration in curriculum development.

5.5 Conclusion

In this dissertation, young GPs' priorities regarding ILO in GP specialty training have been presented for the first time. Comparison with preexisting literature shows an overlap with previous documents published by EURACT and WONCA, except for the novel aspect of doctor's wellbeing. Greater inclusion of leaners into GP curriculum design, as exemplified in this study, may help align ILO with learner needs, and thereby improve the effectiveness and sustainability of CBME in GP.

Furthermore, it was shown that selecting and embedding OLA in a digital learning environment for GP may increase their use. Students' success in digital formative assessment appears to be related to their academic performance. This indicates that learning analytics can enable guidance and monitoring of students' self-directed learning in GP. In the future, learning analytics could be an integral part of e-portfolios and thereby improve feedback in a competency-based curriculum.

6 References

1. Kaduszkiewicz H, Teichert U, van den Bussche H. Ärztemangel in der hausärztlichen Versorgung auf dem Lande und im Öffentlichen Gesundheitsdienst. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. 2018;61(2):187-94.

2. Gesundheitszentren für Deutschland: Wie ein Neustart in der Primärversorgung gelingen kann. 2016 [cited 15.11.2022]. Available from: https://www.bosch-stiftung.de/de/publikation/gesundheitszentren-fuer-deutschland.

3. Windak A, Frese T, Hummers E, Klemenc Ketis Z, Tsukagoshi S, Vilaseca J, et al. Academic general practice/family medicine in times of COVID-19 - Perspective of WONCA Europe. Eur J Gen Pract. 2020;26(1):182-8.

4. Holmboe ES. The transformational path ahead: competency-based medical education in family medicine. Fam Med. 2021;53(7):583-9.

5. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. Med Teach. 2010;32(8):638-45.

6. McGaghie WC, Miller GE, Sajid AW, Telder TV. Competency-based curriculum development on medical education: an introduction. Public Health Pap. 1978(68):11-91.

7. Van Merrienboer JJG. Instructional Design for Competency-based Learning. Interactive Educational Multimedia. 2001:12-26.

8. Long DM. Competency-based Residency Training: The Next Advance in Graduate Medical Education. Academic Medicine. 2000;75(12).

9. Wolcott MD, Quinonez RB, Tittemore AJ, LaGarry-Cahoon A, Zomorodi M, Reside JM, et al. Developing a curriculum blueprint and taxonomy: Process, progress, and pivots. Journal of Dental Education. 2020;84(12):1368-77.

10. Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. Med Teach. 2007;29(7):642-7.

11. Edgar L, McLean S, O. Hogan S, Hamstra S, Holmboe E. Accreditation Council for Graduate Medical Education: The Milestones Guidebook. 2020.

12. Michels NRM, Maagaard R, Buchanan J, Scherpbier N. Educational training requirements for general practice/family medicine specialty training: recommendations for trainees, trainers and training institutions. Educ Prim Care. 2018;29(6):322-6.

13. Konopasek L, Norcini J, Krupat E. Focusing on the formative: building an assessment system aimed at student growth and development. Acad Med. 2016;91(11):1492-7.

14. Lee GB, Chiu AM. Assessment and feedback methods in competency-based medical education. Ann Allergy Asthma Immunol. 2022;128(3):256-62.

15. Shute VJ. Focus on Formative Feedback. Review of Educational Research. 2008;78(1):153-89.

16. Hattie J, Timperley H. The Power of Feedback. Review of Educational Research. 2007;77(1):81-112.

17. Rushton A. Formative assessment: a key to deep learning? Medical Teacher. 2005;27(6):509-13.

18. French JC, Colbert CY, Pien LC, Dannefer EF, Taylor CA. Targeted Feedback in the Milestones Era: Utilization of the Ask-Tell-Ask Feedback Model to Promote Reflection and Self-Assessment. J Surg Educ. 2015;72(6):e274-9.

19. Rose S. Medical Student Education in the Time of COVID-19. JAMA. 2020;323(21):2131-2.

20. Thoma B, Turnquist A, Zaver F, Hall AK, Chan TM. Communication, learning and assessment: Exploring the dimensions of the digital learning environment. Med Teach. 2019;41(4):385-90.

21. Lockyer J, Carraccio C, Chan M-K, Hart D, Smee S, Touchie C, et al. Core principles of assessment in competency-based medical education. Medical Teacher. 2017;39(6):609-16.

22. Saqr M. A literature review of empirical research on learning analytics in medical education. Int J Health Sci (Qassim). 2018;12(2):80-5.

23. Tempelaar DT, Rienties B, Giesbers B. In search for the most informative data for feedback generation: Learning analytics in a data-rich context. Computers in Human Behavior. 2015;47:157-67.

24. Lockyer L, Heathcote E, Dawson S. Informing Pedagogical Action Aligning Learning Analytics With Learning Design. American Behavioral Scientist. 2013;57:1439-59.

25. Danilovich N, Kitto S, Price DW, Campbell C, Hodgson A, Hendry P. Implementing Competency-Based Medical Education in Family Medicine: A Narrative Review of Current Trends in Assessment. Fam Med. 2021(53(1)):9-22.

26. Al-Amri S, Ali Z. Systematic Review of Computer Based Assessments in Medical Education. Saudi J Med Med Sci. 2016;4(2):79-88.

27. Say R, Visentin D, Cummings E, Carr A, King C. Formative online multiple-choice tests in nurse education: An integrative review. Nurse Education in Practice. 2022;58:103262.

28. Wynter L, Burgess A, Kalman E, Heron JE, Bleasel J. Medical students: what educational resources are they using? BMC Med Educ. 2019;19(1):36.

29. Cong X, Zhang Y, Xu H, Liu LM, Zheng M, Xiang RL, et al. The effectiveness of formative assessment in pathophysiology education from students' perspective: a questionnaire study. Adv Physiol Educ. 2020;44(4):726-33.

30. Larsen DP, Butler AC, Roediger HL, 3rd. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. Med Educ. 2009;43(12):1174-81.

31. Rohrer D, Taylor K, Sholar B. Tests enhance the transfer of learning. Journal of experimental psychology Learning, memory, and cognition. 2010;36:233-9.

32. Bringman-Rodenbarger L, Hortsch M. How students choose E-learning resources: The importance of ease, familiarity, and convenience. FASEB Bioadv. 2020;2(5):286-95.

33. Childs S, Blenkinsopp E, Hall A, Walton G. Effective e-learning for health professionals and students—barriers and their solutions. A systematic review of the literature—findings from the HeXL project. Health Information & Libraries Journal. 2005;22:20-32.

34. O'Doherty D, Dromey M, Lougheed J, Hannigan A, Last J, McGrath D. Barriers and solutions to online learning in medical education - an integrative review. BMC Med Educ. 2018;18(1):130.

35. Regmi K, Jones L. A systematic review of the factors – enablers and barriers – affecting e-learning in health sciences education. BMC Medical Education. 2020;20(1):91.

36. Turoff M. The design of a policy Delphi. Technological Forecasting and Social Change. 1970;2(2):149-71.

37. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs. 2000;32(4):1008-15.

38. McKenna HP. The Delphi technique: a worthwhile research approach for nursing? Journal of Advanced Nursing. 1994;19(6):1221-5.

39. Dalkey N, Helmer O. The Delphi method : an experimental study of group opinion. 1969.

40. Bloom BS. Taxonomy of educational objectives: the classification of educational goals. New York: Longmans Green; 1956.

41. de Loe RC. Exploring complex policy questions using the policy Delphi: a multi-round, interactive survey method. Applied Geography. 1995;15(1):53-68.

42. Wijnen-Meijer M, van der Schaaf M, Nillesen K, Harendza S, Ten Cate O. Essential facets of competence that enable trust in graduates: a delphi study among physician educators in the Netherlands. J Grad Med Educ. 2013;5(1):46-53.

43. Junge H, Poppleton A, Sun S, Janos S, Dupont F. What competencies do European General Practice trainees value the most? A prioritisation exercise using a modified Delphi approach. (Submitted). 2023.

44. Junge H, Salzmann A, Schuster K, Volz-Willems S, Jäger J, Dupont F. The use of digital learning analytics in a blended-learning family medicine curriculum at Saarland University. (Submitted). 2023.

45. Wenghofer EF, Kam SM, Timony PE, Strasser R, Sutinen J. Geographic variation in FP and GP scope of practice in Ontario: comparative provincial study. Can Fam Physician. 2018;64(6):e274-e82.

46. Steinhäuser J, Jäkel K, Szecsenyi J, Goetz K, Ledig T, Joos S. Procedures performed in general practice - a cross-sectional study. Gesundheitswesen. 2017;79(12):1004-11.

47. Hafferty FW, Franks R. The hidden curriculum, ethics teaching, and the structure of medical education. Acad Med. 1994;69(11):861-71.

48. Dornan T. Medical education theory and practice. Edinburgh: Churchill Livingston, Elsevier; 2011. 257-64. p.

49. Chan TE, Lockhart JS, Schreiber JB, Kronk R. Determining nurse practitioner core competencies using a Delphi approach. J Am Assoc Nurse Pract. 2020;32(3):200-17.

50. Prasad H, Foong C-C, Hong W-H. Development of a list of core clinical topics suitable for a shared primary care curriculum for medical undergraduates in Malaysia: A Delphi study. Education for Primary Care. 2022;33(5):258-64.

51. Song Y, Shin J, Kim Y, Shim JY. Defining Essential Topics and Procedures for Korean Family Medicine Residency Training. Korean J Fam Med. 2021;42(6):477-82.

52. De Vocht K, Verheyen K, Michels NR. Learning objectives of Belgian general practitioner trainees regarding their hospital training: A qualitative study. Eur J Gen Pract. 2022;28(1):173-81.

53. Derksen F, Olde Hartman T, Lagro-Janssen A, Kramer A. Clinical empathy in GP-training: Experiences and needs among Dutch GP-trainees. "Empathy as an element of personal growth". Patient Educ Couns. 2021;104(12):3016-22.

54. ten Cate O. Summative assessment of medical students in the affective domain. Medical Teacher. 2009;22:40-3.

55. Guldal D, Windak A, Maagaard R, Allen J, Kjaer NK. Educational expectations of GP trainers. A EURACT needs analysis. Eur J Gen Pract. 2012;18(4):233-7.

56. Zarbailov N, Wilm S, Tandeter H, Carelli F, Brekke M. Strengthening general practice/family medicine in Europe-advice from professionals from 30 European countries. BMC Fam Pract. 2017;18(1):80.

57. Scott K, Morris A, Marais B. Medical student use of digital learning resources. Clin Teach. 2018;15(1):29-33.

58. Khalil R, Mansour AE, Fadda WA, Almisnid K, Aldamegh M, Al-Nafeesah A, et al. The sudden transition to synchronized online learning during the COVID-19 pandemic in Saudi Arabia: a qualitative study exploring medical students' perspectives. BMC Med Educ. 2020;20(1):285.

59. Saks K, Leijen Ä. Distinguishing Self-Directed and Self-Regulated Learning and Measuring Them in the E-Learning Context. Procedia - Social and Behavioral Sciences. 2014;112:190-8.

60. Velan GM, Jones P, McNeil HP, Kumar RK. Integrated online formative assessments in the biomedical sciences for medical students: benefits for learning. BMC Medical Education. 2008;8(1):52.

61. Baig M, Gazzaz ZJ, Farouq M. Blended Learning: The impact of blackboard formative assessment on the final marks and students' perception of its effectiveness. Pak J Med Sci. 2020;36(3):327-32.

62. Morrison C, Smith L, Ross L, Butler A, Smith C. Learning through Self-Assessment: Investigating the Relationship between Performance on the NBME® Clinical Science Mastery Series Self-Assessments and Clinical Science Subject Examinations. Medical Science Educator. 2016;26(4):665-72.

63. Costich M, Finkel MA, Friedman S, Catallozzi M, Gordon RJ. Transition-to-residency: pilot innovative, online case-based curriculum for medical students preparing for pediatric internships. Med Educ Online. 2021;26(1):1892569.

64. Roskvist R, Eggleton K, Goodyear-Smith F. Provision of e-learning programmes to replace undergraduate medical students' clinical general practice attachments during COVID-19 stand-down. Education for Primary Care. 2020;31(4):247-54.

65. Johnson C, Shen E, Winn K, Digiacobbe G, Akinola M. Neonatal Resuscitation: A Blended Learning Curriculum for Medical and Physician Assistant Students. MedEdPORTAL. 2020;16:10921.

66. Margolin EJ, Kurtzman JT, Gordon RJ, Anderson CB, Badalato GM. Efficacy of an Online Blended Learning Curriculum to Improve Medical Student Urologic Education. Med Sci Educ. 2021;31(6):2007-15.

67. Chandler P, Sweller J. Cognitive Load Theory and the Format of Instruction. Cognition and Instruction. 1991;8(4):293-332.

68. Bopp C, Salzmann A, Volz-Willems S, Jäger J, Dupont F. The Role of Providing Structure in a Family Medicine Curriculum for Undergraduate Medical Education. Zeitschrift für Allgemeinmedizin. 2022;98 (11):396-401.

69. Prober CG, Khan S. Medical education reimagined: a call to action. Acad Med. 2013;88(10):1407-10.

70. Schuwirth L, Ash J. Assessing tomorrow's learners: in competency-based education only a radically different holistic method of assessment will work. Six things we could forget. Med Teach. 2013;35(7):555-9.

71. Epstein RM. Assessment in Medical Education. New England Journal of Medicine. 2007;356(4):387-96.

72. Kuhn S, Frankenhauser S, Tolks D. [Digital learning and teaching in medical education : Already there or still at the beginning?]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2018;61(2):201-9.

73. Hrynchak P, Takahashi SG, Nayer M. Key-feature questions for assessment of clinical reasoning: a literature review. Med Educ. 2014;48(9):870-83.

74. Siemens G. Learning Analytics: The Emergence of a Discipline. American Behavioral Scientist. 2013;57(10):1380-400.

75. Gil P, Martins S, Moro S, Costa J. A data-driven approach to predict first-year students' academic success in higher education institutions. Education and Information Technologies. 2021;26.

76. Saqr M, Fors U, Tedre M. How learning analytics can early predict under-achieving students in a blended medical education course. Medical Teacher. 2017;39(7):757-67.

77. Ellaway R, Masters K. AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment. Medical Teacher. 2008;30(5):455-73.

78. Challis M. AMEE Medical Education Guide No.11 (revised): Portfolio-based learning and assessment in medical education. Medical Teacher. 1999;21(4):370-86.

79. Van Tartwijk J, Driessen EW. Portfolios for assessment and learning: AMEE Guide no. 45. Med Teach. 2009;31(9):790-801.

80. Dannefer EF, Henson LC. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. Acad Med. 2007;82(5):493-502.

81. Davis MH, Friedman Ben-David M, Harden RM, Howie P, Ker J, McGhee C, et al. Portfolio assessment in medical students' final examinations. Med Teach. 2001;23(4):357-66.

7 Abbreviations

AC	affective competencies
BL	blended learning
CBME	competency-based medical education
CC	cognitive competencies
e.g.	exempli gratia
EURACT	European Academy of Teachers in General Practice
EYFDM	European Young Family Doctors Movement
GP	general practice
GPs	general practitioners
ILO	intended learning outcomes
IMPP	Institut für medizinische und pharmazeutische Prüfungsfragen
LoA	level of agreement
MCQ	multiple choice question
MCQs	multiple choice questions
N	study population size
n	sample size
OLA	online learning activities
p	P-Value
PC	psychomotor competencies
SD	standard deviation
UdS	Saarland University/Universität des Saarlandes
VdGM	Vasco da Gama Movement
WONCA	World Organization of Family Doctors
ρ	Spearman's rho
z.B.	zum Beispiel

8 List of tables

Table 1: Subgroups and examples for the domain of ‘psychomotor competencies’

Table 2: Subgroups and examples for the domain of ‘cognitive competencies’

Table 3: Subgroups and examples for the domain of ‘affective competencies’

Table 4: Psychomotor competencies (PC), cognitive competencies (CC), affective competencies (AC) and their rank, Level of Agreement (LoA) and skewness in Round 1 of the Delphi study. Green color: High or medium LoA. Yellow color: low LoA. Red color: no consensus [43].

Table 5: Psychomotor competencies (PC), cognitive competencies (CC), affective competencies (AC) and their rank, Level of Agreement (LoA) and skewness in Round 2 of the Delphi study. Green color: High or medium LoA. Yellow color: low LoA. Red color: no consensus [43].

Table 6: Descriptive analysis of digital learning behaviour for UdS students in the summer semester 2021 [44].

Table 7: Descriptive analysis of digital learning behaviour for UdS students in the winter semester 2020/21 [44].

Table 8: Correlation matrix: online learning behaviour, formative assessment, and academic performance for UdS students in the summer semester 2021 [44].

Table 9: Correlation matrix: online learning behaviour, formative assessment, and academic performance for UdS students in the winter semester 2020/21 [44].

Table 10 (appendix): Correlation matrix: online learning behaviour and formative assessment for UdS students in the summer semester 2021 [44].

Table 11 (appendix): Correlation matrix: online learning behaviour and formative assessment for UdS students in the winter semester 2020/21 [44].

9 List of figures

Figure 1: Scheme of a ‚curricular blueprint' [5, 9]

Figure 2: European Definition of Family Medicine: Core Competencies and Characteristics [12]

Figure 3: Scheme of the Delphi study

Figure 4: Countries of origin for participants in the Delphi rounds and number of participants per round (Round 1/Round 2). Not on map: Kyrgyzstan (1/0), Malta (1/0), Israel (1/1). Created using www.mapchart.net [43].

10 Publications

Parts of this dissertation were submitted to the peer-reviewed journal ‘Education for Primary Care’ for publication [43]:

Junge H, Poppleton A, Sun S, Janos S, Dupont F. What competencies do European general practice trainees value the most? A prioritisation exercise using a modified Delphi approach (Submitted). 2023.

11 Appendix

		Selected learning cards read	Access times selected cards	Total learning cards read	Question success rate	Total no. of questions answered	Formative assessment score
Selected learning cards read	Spearman's rho	—					
	p-value	—					
Access times selected cards	Spearman's rho	0.525	—				
	p-value	< .001	—				
Total learning cards read	Spearman's rho	0.162	0.629	—			
	p-value	0.110	< .001	—			
Question success rate	Spearman's rho	0.302	0.209	0.013	—		
	p-value	0.003	0.039	0.895	—		
Total no. of questions answered	Spearman's rho	0.359	0.528	0.667	0.131	—	
	p-value	< .001	< .001	< .001	0.197	—	
Formative assessment score	Spearman's rho	0.217	0.281	0.215	0.789	0.314	—
	p-value	0.034	0.005	0.035	< .001	0.002	—

Table 10: Correlation matrix: online learning behaviour and formative assessment scores for UdS students in the summer semester 2021 [44].

		Selected learning cards read	Access times selected cards	Total learning cards read	Question success rate	Total no. of questions answered	Formative assessment score
Selected learning cards read	Spearman's rho	—					
	p-value	—					
Access times selected cards	Spearman's rho	0.632	—				
	p-value	< .001	—				
Total learning cards read	Spearman's rho	0.135	0.550	—			
	p-value	0.255	< .001	—			
Question success rate	Spearman's rho	0.190	0.163	-0.100	—		
	p-value	0.003	0.216	0.450	—		
Total no. of questions answered	Spearman's rho	-0.128	-0.248	-0.368	-0.093	—	
	p-value	0.281	0.035	0.001	0.480	—	
Formative assessment score	Spearman's rho	0.331	0.275	0.042	0.419	-0.109	—
	p-value	0.005	0.020	0.729	0.001	0.361	—

Table 11: Correlations for online learning behaviour and formative assessment scores for UdS students in the winter semester 2020/21 [44].

12 CV

For data protection reasons, the curriculum vitae is not published in the electronic version of the dissertation.

Aus datenschutzrechtlichen Gründen wird der Lebenslauf in der elektronischen Fassung der Dissertation nicht veröffentlicht.

13 Acknowledgements

I would like to show my deep appreciation to my supervisors Prof. Dr. Johannes Jäger and Dr. Fabian Dupont who guided me throughout this project. I also wish to thank GradUS (Support Center for Doctoral Students at Saarland University) for financially supporting my conference visits to Edinburgh and London.