

What is medical education research? An analysis and definition of subjects, objectives and types of research based on articles that have undergone a peer review process

Abstract

Background: Medical education research (MER) seeks to contribute to the scientific knowledge in this area and to the further development of educational practice. However, the lack of relevance of the studies conducted in this field has been criticized for years. The present work therefore aims to clarify the tasks and objectives of MER and the nature of research in this area. To this end, the subjects, objectives, and types of research that are common to MER are analyzed and categories for these are developed.

Method: The categories for *research subjects* and *research objectives* were developed iteratively in three phases with multiple rounds based on samples of peer-reviewed articles. Depending on the round, two to six people were involved in the independent categorization, finding mutual consent, and further development of the categories. At the same time, *research types* were defined for MER.

Results: 169 articles were assessed. Eleven subject categories and eight categories of research objectives were identified, and four types of research were defined as relevant.

Discussion: The categories found for *research subjects* partly coincide with existing category systems but also broaden them. The *research objectives* identified are more specific than they have been before, which limits the scope for interpretation.

Conclusion: The category systems developed can help to define the *subjects* and *objectives* of medical education research more precisely and to differentiate between the *research types* and their significance. In addition, trends and temporary phenomena can also be depicted using the categories found.

Keywords: medical education research, medical education, basic research, applied research, use-inspired basic research

Katrin

Schüttpelz-Brauns¹

Achim Schneider²

Götz Fabry³

Jan Matthes⁴

Monika Himmelbauer⁵

Beatrice Buss⁶

Marianne Giesler⁷

1 Heidelberg University,
Medical Faculty Mannheim,
Mannheim, Germany

2 University of Ulm, Medical
Faculty, Ulm, Germany

3 University Freiburg, Medical
Faculty, Freiburg i. Brsg.,
Germany

4 University of Cologne, Faculty
of Medicine, Cologne,
Germany

5 Medical University of Vienna,
Vienna, Austria

6 Bern University of Applied
Sciences, Department of
Health, Bern, Switzerland

7 Freiburg i. Brsg., Germany

1. Introduction

Research in higher education originated in the United States in the 1920s [1]. In medicine, it became established there from the 1950s under the name *medical education research* [2], [3]. Medical education research was a reaction to various socio-historical factors, such as

“...the increasing importance of scientific research, the availability of funds for MER [medical education research], the explosive growth of medical knowledge, and concerns about accountability for, and control of, medical education.” [4]

For years, international journals have repeatedly criticized the lack of relevance of research in medical education research, which can be summarized as a lack of contribution to science in terms of knowledge gain and a lack of implications for practice. The criticism of the lack of scientific knowledge gain relates, e.g., to irrelevant topics or the lack of generalizability of results, the criticism of the lack of practical implications relates, e.g., to the applicability of the results to practice, but also to the measured outcomes. The criticisms are presented and explained in table 1.

To understand how research projects in medical education research could contribute to scientific knowledge or how to further develop practice, it is worth taking a closer look at medical education research. The first question to ask is what specific topics or subjects are being researched in medical education research (*research subjects*). Also, of relevance is what types of research objectives are pursued (*research objectives*) and what types of research (*research type*) are conducted. These three areas of medical education research are examined in more detail in the following sections.

1.1. Subjects of medical education research

Looking at relevant textbooks, the wide range of *subjects* of medical education research becomes apparent [5], [6], [7], [8]. This diversity is also reflected in articles in which the *subjects* of medical education research studies are systematically analyzed and categorized [9], [10]. These categories are not always congruent. Moreover, since each *research subject* can be viewed or researched from different perspectives, a one-dimensional enumeration of the *research subjects* is difficult to achieve.

Since, to our knowledge, there is no generally accepted classification of *research subjects* in studies of medical education research, a structural model developed in educational psychology [11] could be helpful for future categorization of *subjects* in medical education research. In this model, the *subjects* of educational psychology research are arranged along three dimensions. These are the *functional areas* of research, counselling, prevention, intervention as well as monitoring and evaluation, the *educational career*, which covers the entire lifespan and

thus emphasizes lifelong learning, and the *activity levels* (micro, meso, macro) at which the tasks are to be performed. Based on this model, the *subjects* of medical education research such as student selection, learning, teaching, examination, curriculum could be assigned to the dimension of *functional areas*, which could be scientifically analyzed at different levels (individual, group, organization), but also in relation to the *educational career* (undergraduate education, post-graduate education, further education).

1.2. Research objectives of medical education research

In medical education, there are two frameworks that address corresponding *research objectives* of medical education research. According to Cook et al., studies aim to describe (*description*, focus on observation), justify (*justification*, “How does an intervention work?”) or explain (*clarification*, “How does something work?”, “Why does something work?”) [12]. In their *research compass*, Ringsted et al. summarize the *research objectives* under the categories of *modelling*, *justifying*, *predicting* and *implementing* [13]. However, both papers focus on the type of data collection or study design to answer these questions. How the categories of *research objectives* were developed is left open.

1.3. Types of research in medical education

Looking at the literature on the classification of *research types* in medical education, there are several researchers who place them on a one-dimensional continuum. For example, Albert, who analyzed the discussions in medical education research from a sociological perspective using the field concept of Pierre Bourdieu, found that medical education research is conducted by two interest groups [14]. One group seeks to expand and deepen the knowledge of the research field (epistemic interest). This interest group represents *basic research*. Studies of this *type of research* aim to fundamentally understand general phenomena and their relationships [13], [15], [16]. The aim is to deepen the knowledge and understanding of learning and training within the framework of medical education research [13]. The focus here is on the question “Why does something work?” [13], [16]. The derivation of the research question must therefore be embedded in the existing knowledge on the topic and, if applicable, in a theoretical framework. The other interest group seeks to solve practical problems in medical education with the help of research [14] (application interest). This group represents *applied research*. Studies of this *type of research* aim to solve problems for practice [17]. These can concern both the effectiveness of measures on learning [18], as well as the effect of measures on society or patients [19], [20], [21], [22]. The relevance of research from the perspective of applied researchers is

Table 1: Points of criticism of studies in medical education research

Point of criticism	Aspects	Explanations
Knowledge gain	Irrelevant topics or topics of marginal interest [19], [36], [37], [38]	Topics are not or only marginally relevant for the further development of medical education
	Opportunistically selected study [17], [20], [39]	Conducting studies because the opportunity arises (e.g. when changing teaching formats), regardless of the current state of knowledge
	More description than explanation [9], [12]	Describing new teaching or assessment methods rather than examining general principles of effectiveness
	Lacking generalizability [21], [36], [40]	Generalizability of the results not given
	Lacking conceptual embedding [17], [20], [22], [36], [39], [41]	No reference to the current state of knowledge or theoretical frameworks
Practical implications	Lacking practical applicability of results [36]	No implication for education practice
	Satisfaction instead of patient outcome [20], [22]	Evaluation of effectiveness only at the first level of the Kirkpatrick Model, fourth level desired

primarily concerned with the practical implications of research findings for improving medical education [13]. The one-dimensional view with the extremes of epistemic interest vs. application interest also exists in other research disciplines and has been described in detail by Stokes [23]. However, Stokes points out that the objectives of studies should be considered two dimensionally. These dimensions represent the presence or absence of the *quest for fundamental understanding* and *considerations of use*. In his two-dimensional quadrant model, three different types of research with different objectives can be distinguished [23]. *Pure basic research* is the study of phenomena in a scientific field in order to understand them. It is driven solely by trying to understand without thought of practical use [23]. *Pure applied research* is aimed at an individual, group- or society-related need or application. This is purely for the sake of application with no claim to gain knowledge [23]. The two-dimensional representation generates a third quadrant in which both *quest for fundamental understanding* and *considerations of use* are high. This quadrant represents *use-inspired basic research*. In this type of research, unknown basic principles are tested, and society-related needs are met [23].

Stokes gives concrete examples for three of the quadrants. Although the fourth quadrant is unnamed, it is by no means empty. Stokes assigns research to this quadrant that is not inspired by the goal of understanding nor by the goal of use, but instead systematically investigates certain phenomena [23], [24].

1.4. Aim and research questions

Previous attempts to describe medical education research in more detail on the basis of its *subjects*, *objectives* and *types of research*, in order to ultimately demonstrate the significance of its studies in terms of their contributions to knowledge and implications for practice, has revealed a number of gaps. For example, due to the diversity of topics and subjects, there are no uniform *subject categories* in medical education research. Similarly, there are no

concrete categories of *research objectives* that are or should generally be pursued in medical education research. And even if the assignment of studies in medical education research to Stokes' four possible categories appears to be feasible, there is no precise knowledge of the *type of research* conducted in medical education research. To fill these gaps, this study aims to clarify the *subjects* and *objectives* of medical education research and the *nature of research* in this area. To do this, we need to answer the following three questions:

1. What subjects are researched in medical education?
2. What are the objectives of medical education research?
3. Which types of research listed in Stokes' quadrant model [23] can be derived for medical education research

2. Methods

It can be assumed that peer-reviewed medical education research articles reflect the possible range of *research subjects*, *research objectives* and *types of research* from the perspective of the authors, reviewers and editors. To test this assumption, a method based on document analysis [25] was chosen. In three phases with multiple reviews and samples, selected articles were categorized according to area of *research subject*, *research objective* and *type of research*. In phase 1 (*inductive approach* [26]), categories were developed for these areas to which the studies could be assigned. Phase 2 (*deductive approach* [26]) was used to revise the categories. In phase 3 (*fine-tuning*), the categories and their explanations were refined. Table 2 provides an overview of all phases.

2.1. Samples

In total, there were five samples of articles (SP1 to S5) that differed in terms of source as well as inclusion and exclusion criteria.

Table 2: Overview of the phases and rounds of the study

Round	Sample	N _{Article}	N _{Rater}	Aim of the round
Phase 1: Inductive approach				
R1	S1	30	2	Initial categories for <i>research objectives</i> and classification of studies into <i>research types</i>
R2	S2	50	5	Review of the categories for <i>subjects</i> and <i>research objectives</i> , as well as classification of <i>research types</i>
Phase 2: Deductive approach				
R3	S3	30	2	Further development of the categories for <i>research subjects</i> and <i>research objectives</i> , as well as classification of <i>research types</i> including review of the categories
R4	S3	12	2	
R5	S4	15	2	
R6	S5	60	5	Review of the categories for <i>research subjects</i> and <i>research objectives</i> , as well as classification of <i>research types</i> in the enlarged group
R7	S1-S5	50	5	Further development of the categories for <i>research subjects</i> and <i>research objectives</i> and classification of <i>research types</i> . Review of the categories for <i>research subjects</i> and <i>research objectives</i> , as well as classification of <i>research types</i>
R8	S1-S5	20	6	
Phase 3: Fine-tuning				
R9	S1-S5	20	3	Revision and refining of the categories for <i>research subjects</i> , <i>research objectives</i> , and <i>research types</i> with subsequent review
R10	S1-S5	20	3	
R11	S1-S5	20	3	

R: round; N_{Article}: Number of articles used; N_{Rater}: Number of raters; S1-S5: random selection from the population "all articles from the previous rounds that were published in a journal from medical education, excluding duplicates"

The samples were drawn independently of each other to be able to cover as broad a field as possible when developing categories or reviewing them. Towards the end of the second phase, it became apparent that a further sampling of additional articles would not have led to any new findings, so the articles from S1 to S5 were transferred to a single file. Duplicate articles were identified using the DOI and removed. Articles were then randomly selected from the resulting file, excluding studies from purely medical journals. A detailed description of the sample selection can be found in attachment 1.

2.2. Procedure

The categories were developed iteratively (see figure 1) over several rounds (see table 2). The categories for *research subjects* and *research objectives* were developed in terms of content, while the categories for the *types of research* according to Stokes [23] only had to be adapted for medical education.

The professional background of four researchers involved in this study is psychology (MG, MH, AS, KSB), two researchers studied medicine (GF, JM) and one physiotherapy (BB).

2.2.1. Phase 1: Inductive approach (round 1 – round 2)

Round 1 aimed to find initial categories for *research objectives* and to assign articles to the *research types* according to Stokes [23] (see table 2). For this purpose, two authors (MG, KSB) independently paraphrased the

research objective of the article, the epistemic interest and the application interest. The *research type* was derived from the combination of epistemic and application interest. Discrepancies in the assignment of *research objectives* or *research types* addressed were discussed until agreement was reached. The inductive method was chosen to develop *research objectives* independently of the categories of Cook [12] and Ringsted [13] and, if necessary, to validate these with different methods in the sense of triangulation.

The aim of round 2 was to review the categories for *research subjects*, *research objectives* and *research types* by five independent researchers (BB, MH, JM, AS, KSB). The categories for *subjects* and *objectives* as well as the classification into *research types* were predetermined. The structural model of educational psychology described above [11] formed the basis for the development of the initial categories for *subjects* of medical education research. Each dimension of this model was assigned to the study subjects commonly used in medical education research (see table 3). The categories for the *research objectives* were derived from the results of round 1. Since it was assumed that the *research objectives* could be assigned to the categories of the *research types*, the respective assignment of the *research objectives* to the *research types* was predetermined.

For reasons of research economy, the allocation of articles to be categorized was based on a balanced plan (see attachment 2).

Mismatches in the assignment to the categories were identified. If possible, the categories were subsequently

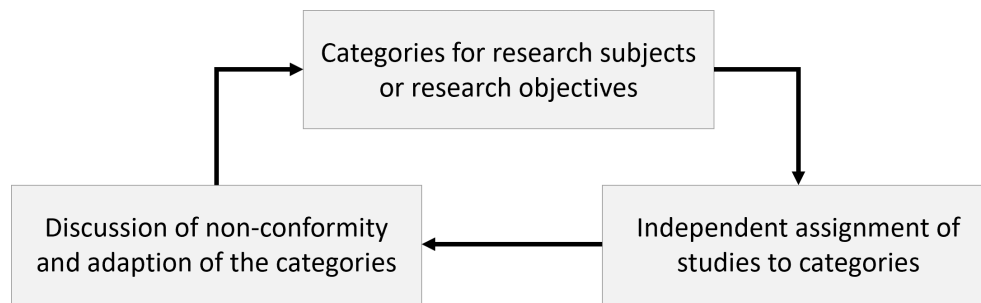


Figure 1: An Iterative approach for the development of categories for subjects and objectives of medical education research

Table 3: Initial categories for subjects in medical education research

Dimension	Initial categories
Functional areas	Student selection, learning, teaching, examination, curriculum, supervision/leadership, management
Educational career	Undergraduate education, post-graduate education, further education
Activity levels	Individuals, groups, organizations

reassigned, considering the comments on each *subject* or *objective*.

2.2.2. Phase 2: Deductive approach (round 3 – round 8)

Round 3 to round 5 were used to further develop the categories for *research subjects*, *research objectives* and classification into *research types* (see table 2). The categories were predefined and derived from the results of the respective previous round. Two researchers (MG, KSB) independently assigned the articles to the categories. Non-matches between the categories were discussed, categories were renamed and reorganized if necessary, and explanations were added to each category.

The aim of round 6 was to review the categories agreed in round 5, including the explanations for *subjects* and *objectives* as well as the classification of *research types*. For this purpose, five researchers (MG, MH, JM, AS, KSB) independently assessed the articles. Again, for research economic reasons, a balanced plan was used to assign the articles. This plan is equivalent to the one in round 2, only that six articles were assessed in each cell instead of five.

In both round 7 and round 8, the categories were discussed and adjusted by MG, MH, JM, AS and KSB if they did not match. Prior to round 8, MG and GF respectively analyzed from a psychological and medical point of view, which articles were considered critical, i.e. those articles were reviewed that had an overlap of several *types of research*. The focus was on *research objectives* and their relationship to *research types*.

2.2.3. Phase 3: Fine-tuning (round 9 – round 11)

In the *fine-tuning phase*, the categories and explanations for *research subjects*, *research objectives* and *research types* were refined (see table 2). Here, too, several rounds

took place with discussion of non-agreement, revision resp. refining of the categories and independent assessment, each with three researchers (MG, AS, KSB).

2.3. Analysis of rater agreement

Kappa measures were calculated to determine rater agreement, which are to be regarded as adjusted for chance [27], [28]. Cohen's kappa (κ) was calculated to determine the agreement of two raters and Fleiss κ to determine the agreement of more than 2 raters.

In phases 1 and 2, Cohen's κ was calculated for the assessments of the ten and 18 pairs evaluated respectively. These were calculated for all categories of the *research subjects*, *research objectives* and *research types*. In a second step, the median of these values was determined for all areas as an estimate of the average agreement between all raters, as the median in this case is the measure that best characterizes the entire sample [27]. In phase 1, the results section does not state the significance, as the κ values determined are based on the assessments of only five articles.

In phase 3, Fleiss κ was calculated, as the given items had to be assessed by three researchers. This measure can be used to determine not only the rater agreement of the entire category system, but also for each single category.

According to Landis and Koch, the following values are used to interpret the degree of agreement: $\kappa < 0.00$ as poor, 0.00-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial and 0.81-1.00 as almost perfect [28]. In the assessments with three people, the assessment rounds were conducted until sufficient assessment agreement was achieved.

Table 4: Description of the samples

Sample	Drawn on	Period of publications	Origin of the articles N (%)	Article Type Review N (%)
S1	12-02-2021	2021 ¹	Med Educ J = 30 ² (100%)	4 (13%)
S2	12-15-2021	2011-2021 Q ₂₅ = 2016 Q ₅₀ = 2018 Q ₇₅ = 2019	Med Educ J = 37 (74%) Med J = 13 (26%)	16 (32%)
S3	12-15-2021	2014-2019 Q ₂₅ = 2014 Q ₅₀ = 2017 Q ₇₅ = 2019	Med Educ J = 41 (98%) Med J = 1 (2%) ³	8 (19%)
S4	10-30-2022	2005-2022 Q ₂₅ = 2012 Q ₅₀ = 2018 Q ₇₅ = 2020	Med Educ J = 15 (100%)	5 (33%)
S5	02-13-2023	2003-2022 Q ₂₅ = 2014 Q ₅₀ = 2017 Q ₇₅ = 2021	Med Educ J = 60 (100%)	6 (10%)

S: Sample, Q₂₅: 25% percentile, Q₅₀=median, Q₇₅=75% percentile; Med Educ J: journal from medical education, Med J: medical journal;

¹ 7 of these available online in advance and not finally published until 2022;

² of which 17 articles (57%) from German-speaking countries;

³ was only identified retrospectively

3. Results

3.1. Sample characteristics

In total, 197 articles were selected. Of these, 28 were included in two samples. Therefore, a total of 169 articles were assessed. During the assessment process, two articles were identified that did not meet the criteria for a study: One article presented a research method and its application in medical education. The other article was a letter to the editor. Both articles were excluded from further analysis.

Table 4 provides information on the articles in each sample.

3.2. Development of the categories

As an iterative approach was adopted in this study, the results of each phase are not only reported in a descriptive manner, as is usual in scientific studies, but are also used to justify the further procedure.

At the end of phase 1, the following rater agreements were identified for ten rater pairs:

- *Subject areas*
 - *Functional areas category*: $\kappa = -0.111-0.737$; $\kappa_{\text{Md}} = 0.143$
 - *Educational career category*: $\kappa = -0.250-1.000$; $\kappa_{\text{Md}} = 0.302$
 - *Activity levels category*: $\kappa = -0.071-0.412$; $\kappa_{\text{Md}} = 0.185$
- *Research objectives*: $\kappa = -0.087-0.800$; $\kappa_{\text{Md}} = 0.243$
- *Research types*: $\kappa = -0.364-0.444$; $\kappa_{\text{Md}} = 0.000$.

During the inductive phase (phase 1), it became apparent that the *research subjects* assigned to the dimensions

of the structural model of educational research (*functional areas, educational career, activity levels*) sometimes overlapped. For example, the *subjects* teaching, learning and curriculum could not always be clearly distinguished from one another. Additionally, some *subjects* could not be clearly assigned to the dimensions of *activity level* and *educational career*. Based on these overlaps, which may have partly caused the low level of rater-agreements, the concept of three dimensions was abandoned and all *research subjects* were grouped into one subject area. The categories for both the *research subjects* and the *research objectives* were gradually adapted in the consensus procedure in the rounds of phase 2. This concerned the categories themselves as well as the explanations for them. Reaching a consensus proved to be particularly difficult, as the researchers of this study had different ideas about science due to their different professional socialization. Additionally, there was a lack of clarity in the terms as well as missing and overlapping categories. It also became apparent that the dimensions *quest for fundamental understanding* and *considerations of use* in the quadrant model according to Stokes [23] can be interpreted differently. With the help of additional literature [23], [29], [30], [31], [32] and the discussion of specific studies, the *research types* were more clearly distinguished from one another.

For the final round of phase 2, the following Cohen's Kappa were calculated for 18 rater pairs:

- *Research subjects*: $\kappa = -0.053-0.722$; $\kappa_{\text{Md}} = 0.214$
- *Research objective*: $\kappa = -0.111-0.643$; $\kappa_{\text{Md}} = 0.268$
- *Research type*: $\kappa = -0.176-1.000$; $\kappa_{\text{Md}} = 0.354$

While fine-tuning, the fourth quadrant was renamed *background research* as suggested by Bush [33]. This

change was accompanied by higher level of rater agreement. At the end of this phase, the rater agreements were as follows: for *research subject* $\kappa_{\text{Fleiss}}=0.63$, $p<0.001$, for *research objective* $\kappa_{\text{Fleiss}}=0.52$, $p<0.001$ and for *research type* $\kappa_{\text{Fleiss}}=0.58$, $p<0.001$. The results for each category are listed in attachment 3, tables A1-A3.

3.3. Subjects, objectives and types of research in medical education research

After completion of the three phases, eleven *research subject* categories, eight categories of *research objectives* and explanations of the *research types* are now available (see table 5, table 6 and table 7).

4. Discussion

The starting point for this study is the often-lamented lack of relevance of medical education research: Medical education research is often criticized for contributing little to the advancement of knowledge in this field or for producing findings that have little practical implication.

To better understand how research projects in medical education research can contribute to science in terms of gaining knowledge or further developing practice, the following questions were asked in this study:

1. What subjects are researched in medical education?
2. What are the objectives of medical education research?
3. Which types of research listed in Stokes' quadrant model [23] can be derived for medical education research?

To answer these questions, research articles that have undergone peer-review were used and analyzed accordingly. This was done in an iterative process in which initial categories of each area were reviewed and adjusted. This process of independently reviewing and adjusting categories alternated until a satisfactory rater agreement was achieved.

The categories found for the *research subjects* partly correspond to previous categories. For example, Raes et al. [9] use similar categories with a slightly different sorting, but in our case, the teaching-learning methods of Raes et al. are divided into teaching methods and traits, motivation and behavior of learners and teachers. We were able to replicate all, but two of Raes et al.'s seven *research subjects*. In the present categorization, the *research subject* faculty development can be assigned to the category's curriculum, quality criteria and indicators or positions, functions, roles, careers, depending on the topic. Correspondingly, the subject of social competence can be assigned to the category traits, motivation and behavior of students and teachers. Compared to the *subject categories* described by Rotgans [10], no matches could be found, as the categories we have developed are formulated in a more general way. For example, the *research subjects* listed by Rotgans, which are issues in

student assessment, objective structured clinical examination, clinical competence assessment, can be assigned to the category with the heading of measuring instruments. The category system developed for *research subjects* in the present study therefore fits in well into previous categorization attempts but also broadens them and has the advantage that new *research subjects* can also be included in this categorization.

As far as the *research objectives* of medical education research are concerned, previous classifications leave room for interpretation. Both Cook [12] and Ringsted [13] list categories for *research objectives*. However, they neither explain how these *research objectives* are arrived at nor what they actually mean. Some of the studies we examined can be assigned to Cook's [12] category *description*. For example, there are studies in which phenomena are described that cannot be explained by current theories. These studies provide a starting point for the further development and empirical testing of theories. On the other hand, other studies, which, for example, describe the spread of certain teaching or examination formats without any theoretical or practical implications, would also fall into the description category according to Cook et al.'s categorization. In the present categorization for *research objectives*, however, such studies are assigned to the category of *presentation of a status quo*.

In summary, the *research objectives* we have identified can be roughly categorized according to previous classifications. However, they are formulated so specifically that they are more likely to guarantee a clear classification, as the scope for interpretation has been largely restricted. Furthermore, additional *research objectives* were identified which relate to the (further) development and verification of measurement instruments and data analysis methods as well as the development of consensus-based, normative guidelines and recommendations.

Regarding the *types of research* considered relevant for medical education research, three *types of research* were derived in our study: *use-inspired basic research*, *pure applied research* and *background research*. However, a category of *pure basic research* could not be identified. Referring to the requirements of Eva and Ringsted, this result appears to be partially consistent, because according to their explanations, medical education research should relate to *use-inspired basic research*, i.e. address both *quest for fundamental understanding* and *considerations of use*. Ringsted writes: "Research in medical education seeks to deepen the knowledge and understanding of learning, teaching and education. It is neither about solving concrete, local problems nor about providing general, universal solutions" ([13], p.695). Eva is primarily concerned with understanding phenomena that are of practical relevance ("*identify phenomena of interest (and practical relevance)*")... "*imperative of understanding*", [16], p.295). To understand these practically relevant phenomena, we need to examine the processes of learning, teaching and education. Since these processes are not directly observable, testable theories that can explain these processes come to the fore. However,

Table 5: Subjects of medical education research incl. explanations

Research Subject	Explanations
Traits, motivation and behavior of learners and teachers	Traits, needs, motives, interests, (learning) achievement, attitudes, and behavior of students/teachers including gender differences, and how learners use teaching/learning opportunities; the focus is on the learning process. <i>Examples: clinical decision making, scientific thinking, resilience, emotional intelligence, professional identity development, reflection</i>
Theories/ models, constructs	Description or explanation of reality with the help of theories, models, and/or constructs; a theory is a system of scientifically grounded statements to explain phenomena, (natural) processes and the underlying principles (<i>examples: Self-Determination Theory by Ryan and Deci, Social Cognitive Learning Theory by to Bandura</i>). A model represents a simplified image of reality or offers fundamental approaches to problem-solving (<i>examples: Miller pyramid, Kern cycle</i>). A construct describes traits or states that are not directly observable (<i>examples: intelligence, motivation</i>).
Teaching method	Didactic actions and measures that are deliberately, systematically, and consciously employed to trigger or influence learning processes [43] <i>Examples: Bedside teaching, simulation teaching, individualized feedback, team-based learning, peer-assisted learning</i>
Curriculum; medical education (at meta-level)	A curriculum is based on teaching objectives and the course of the teaching and learning process or the study program. In particular, it includes statements about the framework conditions for learning [44]. This definition also includes the implicit teaching of learning content and forms, as well as effects on socialization beyond curricula (hidden curriculum) that are not officially provided for. A curriculum can refer to an entire degree program, but also to related course offerings such as electives. This category also includes the consideration of medical education at the meta-level, such as a country's education system. <i>Examples: Master's program, model study program, science curriculum, doctoral study program, hidden curriculum</i>
Other interventions	Interventions designed to support or promote physical or mental well-being; measures designed to lead to an improvement in quality <i>Examples: Mindfulness training</i>
Measuring instruments (incl. summative assessment), formative assessment as a measuring instrument	Means of converting observed data into measured values; this applies especially to measuring instruments for student selection (e.g. <i>HAM-Nat</i>), for performance assessment (examinations, e.g. <i>OSCE, Script Concordance Test</i>) or for use in research (e.g. <i>Approaches to Learning and Studying Inventory, ALSI</i>)
Data analysis methods	Statistical methods for generating quantitative parameters (e.g. <i>Bayesian statistics, cumulative deviation method</i>) or methods for generating related qualitative information (e.g. <i>thematic content analysis according to Braun & Clark</i>)
Quality criteria and indicators	Criteria and indicators with which the quality of teaching formats and training measures can be ensured or assessed <i>Example: Accreditation standards, quality criteria for teaching, quality criteria for medical education research studies</i>
Positions, functions, roles, career	Studies that deal with professional topics, such as <i>women in medicine, career paths of nurses, leadership competencies</i> . Position describes the formal place (rights and duties) that someone holds in a system, in an organization (e.g. <i>medical specialist</i>). Function describes the purpose, and content-related tasks (area of responsibility = tasks and responsibilities) associated with the position (e.g. <i>specialist in surgery</i>). Roles are the personal counterpart to the function. They describe the behavioral expectations directed at the functional holders, e.g. surgical specialist.
Medical education research	Studies in this area aim to assess subjects of medical education research, determine the quality of the research work and learn their methods in order to improve or broaden them as needed. <i>Examples: Status quo of medical education research in Europe, overview of the methods of medical education research</i>
Learning environments, learning conditions, learning contexts	Structural and personal contextual factors for learning, such as <i>digital learning environments, building architecture, deployment abroad, accessibility and availability of teachers, usability of digital tools</i>

Table 6: Objectives of medical education research including explanations and examples

Aim of the research	Explanations
Investigation of traits, motivation, attitudes and behavior (and their changes) of students and teachers, or a value-adding summary of findings in this area	How do students/teachers respond to their environment, to changes in their environment, to demands, etc. or how do they perceive these? How do students/teachers behave as group members? How do students learn? How do traits develop (no interventions)?
(Further) development and / or empirical testing of a theory, a model, a construct or a value-adding summary of findings in this area	How can observed correlations be explained theoretically? Can theoretically postulated relationships be empirically proven? Does a model need to be revised? How can traits and states that are not directly observable be adequately described (construct)? How do stakeholders understand a construct?
Empirical review of teaching methods, teaching materials or other measures relating to learning performance, motivation, satisfaction, attitudes, etc. or a value-adding summary of findings in this area	Is a teaching material, a teaching format or a measure effective in terms of defined outcomes? How do students perceive teaching formats, etc.?
Development and verification of measurement instruments for student selection, for assessment purposes (examination), for research, etc. or a value-adding summary of findings in this area	Which construct should be measured? How is the construct operationalized? How reliable is the instrument measuring what it is intended to measure (reliability)? And does the instrument measure the construct it is intended to measure (validity)? How well do additional test quality criteria, such as test fairness, practicability, etc. perform? Which questionnaires are used in which areas? Are there any differences regarding the target group(s)?
(Further) development and / or review of research approaches, research methodology and data analysis procedures or a value-adding summary of findings in this area	Which research methods are adequate to answer questions in medical education research? How reliable and valid are the results from a newly developed or further developed data analysis method?
Investigation of conditions for the use of teaching methods, learning materials and measures (e.g. as part of problem and feasibility analyses)	Under what conditions are given teaching methods, teaching materials, or other curricular activities suitable for use? Are the measures effective or not? The conditions can be examined, among other things, as follows: (1) Implementation methods of teaching formats incl. conducive and obstructive conditions. (2) Investigating the causes of problems (problems are already known in contrast to the status quo category, see below). (3) Examine the technical/economic feasibility of measures, etc., e.g. feasibility studies.
Development of consensus-based, normative guidelines and recommendations	What learning objectives, teaching content, etc. should be taught in a curriculum? What competencies should teachers have?
Presentation of a status quo	What information can be gained by systematically collecting data on a topic? The aim here is to obtain an overview of a topic, e.g. on the implementation of political guidelines or to be able to identify potential research gaps or practical problems as a result, even if it is not yet clear whether there are any at all.

Table 7: Possible types of research in medical education including explanations

Research type	Explanations
Pure basic research	The sole aim of this research is to gain scientific knowledge for the discipline in terms of understanding general phenomena, correlations or so-called laws. <i>Pure basic research</i> leads to findings for which there is (yet) no direct application (cf. [33] p. 89). The focus on <i>pure basic research</i> is on theories (cf. [45], p. 90).
Use-inspired basic research	Research with the aim of solving a practical problem or examining practical applications, but at the same time striving for scientific knowledge for the discipline. This means that the findings from the study can be used for the further development of theories (cf. [29]).
Pure applied research	Research aimed at finding solutions to practical problems, improving technologies, etc. The focus here is on practically relevant findings that are more tailored to specific practical problems and less generalizable (cf. [45], p. 187). Applied objectives can be the verification of the effect/usefulness/effectiveness of a measure or the test-statistical verification of an instrument.
Background research	Scientific collection and evaluation of data to obtain an overview of a topic or partial aspects of a topic, e.g. conducting inventories (cf. [23], p. 74).

these theories are the subject of *pure basic research* [34]. In medical education research, however, *pure basic research* is still carried out far too rarely. In their systematic review, Tolsgaard et al. assigned only 2% of randomized studies to the Bohr quadrant (= *pure basic research*) [35]. The rarity of the occurrence of such studies may therefore be one reason why no studies were found in the present investigation that can be assigned to *pure basic research*. Due to the importance of this category, as described above, the *research types* were supplemented with the category of *pure basic research*.

4.1. Strengths and weaknesses of the study

The strength of our study lies in the fact that previous classifications of *research objectives* have a strong focus on the research methodology. Yet the choice of research methodology should only take place after a relevant research question has been identified. However, research methodology did not play a role in our discussions and classifications, so we were able to identify the *research objectives* regardless of the choice of the “right” method. It should be emphasized that peer-reviewed articles, which were assumed to be representative of relevant research, were subject to a procedure based on document analysis. An iterative approach was used to develop the categories and their explanations, alternating between phases of finding mutual consent and verification of agreement. Additionally, researchers from different knowledge cultures were involved.

Through the iterative process of sampling, categorization, finding mutual consent, checking of rater agreement, revision of the categorization on further articles, etc., it was possible to develop categories that also withstand trends and time-limited topics. Thus, topics such as artificial intelligence or the COVID-19 pandemic are not literally listed in the categories, but are nevertheless implicitly present, e.g. as teaching or examination format or change in student motivation.

The study also has some weaknesses. Particularly in the *fine-tuning phase* (phase 3), some poorly designed studies complicated the categorization process because, for example, *research subjects* and *research objectives* were not explicitly named or were not stringently addressed in different parts of the articles (e.g. abstract vs. full text). These inconsistencies presumably had a negative effect on the rater agreements. Another problem was the different use of the terms *theory* and *model* in some articles. On the one hand, these terms were used to describe natural processes, which were then tested using empirical data (epistemic meaning). On the other hand, the terms *theory* and *model* were used without claiming to explain or describe behavior and experience. Instead, the aim of these studies was to agree on uniform guidelines (normative meaning). The classification of *research types* was also sometimes affected by shortcomings or ambiguities in the articles. Sometimes *research objectives* were described that could have been assigned to *use-inspired*

basic research but could not be because of the lack of theoretical reference. In summary, such flaws can limit the extent of rater agreements.

Another possible weakness of the present study is that some categories were chosen less frequently to categorize the articles. This could be because studies that fit into these categories are actually less common. Furthermore, it cannot be ruled out that our explanations of these categories are still too imprecise. Our approach was focused on the development of the categories and their explanations, not on determining and considering the actual occurrence of the studies in the individual categories.

5. Conclusion

Our category systems can help define *research subjects* and *research objectives* of medical education research more precisely as well as better distinguish between the *research types* and their significance. This should make it possible to formulate relevant research questions and choose an approach in which explanatory theories play a role or in which the (general) problem of practice is in the foreground, depending on the type of research in question. The next step is to select the research methodology in such a way that the research question can be answered with the help of the data collected and analyzed. Additionally, the terms used, and in particular their meaning, should be discussed and agreed upon, so that those involved in future studies mean the same thing and their work can contribute more effectively to the gain of knowledge in medical education and ultimately to the further development of practice.

Acknowledgements

We would like to thank Olivia Steiger, cand. med. for her contributions to the literature.

Authors' ORCIDs

- Katrin Schüttpelz-Brauns: [0000-0001-9004-0724]
- Achim Schneider: [0000-0002-8602-8535]
- Götz Fabry: [0000-0002-5393-606X]
- Jan Matthes: [0000-0003-2754-1555]
- Monika Himmelbauer: [0000-0001-5516-1993]
- Beatrice Buss: [0000-0001-7270-382X]
- Marianne Giesler: [0000-0001-9384-2343]

Competing interests

The authors declare that they have no competing interests.

Attachments

Available from <https://doi.org/10.3205/zma001806>

1. Attachment_1.pdf (135 KB)
Description of the sample selection
2. Attachment_2.pdf (115 KB)
Balanced plan of 50 articles that were used in phase 1 in the second round to review the categories found
3. Attachment_3.pdf (152 KB)
Assessment agreements at the end of the study

References

1. McKeachie WJ. Research on College Teaching: The Historical Background. *J Educ Psychol.* 1990;82(2):189-200. DOI: 10.1037/0022-0663.82.2.189
2. Norman G. Fifty years of medical education research: waves of migration. *Med Educ.* 2011;45(8):785-791. DOI: 10.1111/j.1365-2923.2010.03921.x
3. Schmidt HH, Mamede S. How cognitive psychology changed the face of medical education research. *Adv Health Sci Educ Theory Pract.* 2020;25(5):1025-1043. DOI: 10.1007/s10459-020-10011-0
4. Kuper A, Albert M, Hodges BD. The origins of the field of medical education research. *Acad Med.* 2010;85(8):1347-1353. DOI: 10.1097/ACM.0b013e3181dce9a7
5. Swanwick T, Forrest K, O'Brien BC, editors. *Understanding Medical Education: Evidence, Theory, and Practice*, 3rd ed. Hoboken (NJ): Wiley Blackwell; 2018. DOI: 10.1002/9781119373780
6. Dent J, Harden R, Hunt D, editors. *A practical guide for medical teachers*. 6th ed. London, UK: Elsevier; 2021.
7. Norman GR, Vleuten CP, Newble DI, Dolmans DH, Mann KV, Rothman A, Curry L, editors. *International Handbook of Research in Medical Education*. Dordrecht: Springer; 2012.
8. Fabry G. *Medizindidaktik. Für eine kompetenzorientierte, praxisrelevante und wissenschaftlich fundierte Ausbildung*. 2. vollst. überarb. Aufl. Bern: Hogrefe; 2023. DOI: 10.1024/85852-000
9. Raes P, Bauer D, Schöppe F, Fischer MR. The active participation of German-speaking countries in conferences of the Association for Medical Education in Europe (AMEE) between 2005 and 2013: A reflection of the development of medical education research? *GMS Z Med Ausbild.* 2014;31(3):Doc28. DOI: 10.3205/zma000920
10. Rotgans JI. The themes, institutions, and people of medical education research 1988-2010. *Adv Health Sci Educ Theory Pract.* 2012;17(4):515-527. DOI: 10.1007/s10459-011-9328-x
11. Spiel C, Wagner P, Götz T, Lüftenegger M, Schober B. *Bildungspsychologie: Konzeption, Strukturmodell, Stellenwert und Perspektiven*. In: Spiel C, Götz T, Wagner P, Lüftenegger M, Schober B, editors. *Bildungspsychologie – ein Lehrbuch*. 2. vollst. überarb. Aufl. Göttingen: Hogrefe; 2022. p.11-30. DOI: 10.1026/03108-000
12. Cook DA, Bordage G, Schmidt HG. Description, justification and clarification: a framework for classifying the purposes of research in medical education. *Med Educ.* 2008;42(2):128-133. DOI: 10.1111/j.1365-2923.2007.02974.x
13. Ringsted C, Hodges B, Scherpbier A. 'The research compass': an introduction to research in medical education: AMEE Guide no. 56. *Med Teach.* 2011;33(9):695-709. DOI: 10.3109/0142159X.2011.595436
14. Albert M. Understanding the debate on medical education research: a sociological perspective. *Acad Med.* 2004;79(10):948-954. DOI: 10.1097/00001888-200410000-00009
15. Eva K. Broadening the debate about quality in medical education research. *Med Educ.* 2009;43(4):294-296. DOI: 10.1111/j.1365-2923.2009.03342.x
16. Eva K. Publishing during COVID-19: Lessons for health education research. *Med Educ.* 2021;55(3):278-280. DOI: 10.1111/medu.14450
17. Albert M, Hodges B, Regehr G. Research in medical education: balancing service and science. *Adv Health Sci Educ Theory Pract.* 2007;12(1):103-115. DOI: 10.1007/s10459-006-9026-2
18. Regehr G. It's NOT rocket science: rethinking our metaphors for research in health professions education. *Med Educ.* 2010;44(1):31-39. DOI: 10.1111/j.1365-2923.2009.03418.x
19. Whitcomb ME. Research in medical education: what do we know about the link between what doctors are taught and what they do? *Acad Med.* 2002;77(11):1067-1068. DOI: 10.1097/00001888-200211000-00001
20. Gruppen LD. Improving medical education research. *Teach Learn Med.* 2007;19(4):331-335. DOI: 10.1080/10401330701542370
21. Leif M, Semarad N, Ganesan V, Selting K, Burr J, Svec A, Clements P, Talmon G. The Quality Of Evidence In Preclinical Medical Education Literature: A Systematic Review. *Adv Med Educ Pract.* 2019;10:925-933. DOI: 10.2147/AMEPS212858
22. Colliver JA, McGaghie WC. The reputation of medical education research: quasi-experimentation and unresolved threats to validity. *Teach Learn Med.* 2008;20(2):101-103. DOI: 10.1080/10401330801989497
23. Stokes DE. *Pasteur's Quadrant. Basic Science and Technological Innovation*. Washington, D.C.: Brookings Institution Press; 1997.
24. Jovanović M. *Fußspuren in der Publikationslandschaft: Einordnung wissenschaftlicher Themen und Technologien in grundlagen- und anwendungsorientierte Forschung mithilfe bibliometrischer Methoden*. Stuttgart: Frauenhofer-Verlag; 2011.
25. Hoffmann N. *Dokumentenanalyse in der Bildungs- und Sozialforschung. Überblick und Einführung*. Weinheim: Beltz Juventa; 2017.
26. Flick U. *Qualitative Forschung. Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften*. 5. Aufl. Reinbek bei Hamburg: Rowohlt; 2000.
27. Wirtz M, Caspar F. *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe; 2002.
28. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
29. Brüggemann A, Bromme R. *Anwendungsorientierte Grundlagenforschung in der Psychologie: Sicherung von Qualität und Chancen in den Beurteilungs- und Entscheidungsprozessen der DFG. Psycholog Rundschau.* 2006;57(2):112-118.
30. Döring N, Bortz J. *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 5. vollst. überarb. aktual. und erweit. Aufl. Berlin: Springer; 2016. DOI: 10.1007/978-3-642-41089-5

31. Reinders H, Gräsel C, Ditton H. Praxisbezug Empirischer Bildungsforschung. In: Reinders H, Ditton H, Gräsel C, Gniewosz B, editors. Empirische Bildungsforschung. Gegenstandsbereiche. Wiesbaden: VS Verl. für Sozialwiss. 2011. p.221-233. DOI: 10.1007/978-3-531-93021-3_19
32. Fischer F, Waibel M, Wecker C. Nutzenorientierte Grundlagenforschung im Bildungsbereich. Zeitschr Erzwiss. 2005;8(3):427-442.
33. Bush V. Science: The Endless Frontier; A report to the President on a Program for Postwar Scientific Research. Reprint anlässlich des 75. Jahrestages. National Science Foundation; 2020. Zugänglich unter/available from: https://www.nsf.gov/about/history/EndlessFrontier_w.pdf
34. Ellaway RH, Hecker KG. What role does basic research have in an applied field? Adv Health Sci Educ Theory Pract. 2022;27(2):289-292. DOI: 10.1007/s10459-022-10117-7
35. Tolsgaard MG. Medical education research in obstetrics and gynecology. Am J Obst Gyn. 2019;220(1):121. DOI: 10.1016/j.ajog.2018.09.013
36. Meyer HS, Durning SJ, Sklar DP, Maggio LA. Making the first cut: an analysis of academic medicine editors' reasons for not sending manuscripts out for external peer review. Acad Med. 2018;93(3):464-470. DOI: 10.1097/ACM.0000000000001860
37. Prystowsky JB, Bordage G. An outcomes research perspective on medical education: the predominance of trainee assessment and satisfaction. Med Educ. 2001;35(4):331-336. DOI: 10.1046/j.1365-2923.2001.00910.x
38. Biesta GJJ, van Braak M. Beyond the medical model: Thinking differently about medical education and medical education research. Teach Learn Med. 2020;32(4):449-456. DOI: 10.1080/10401334.2020.1798240
39. Norman G. Editorial - how bad is medical education research anyway? Adv Health Sci Educ Theory Pract. 2007;12(1):1-5. DOI: 10.1007/s10459-006-9047-x
40. Lurie SJ. Raising the passing grade for studies of medical education. JAMA. 2003;290(9):1210-1212. DOI: 10.1001/jama.290.9.1210
41. Bolander Laksov K, Dornan T, Teunissen PW. Making theory explicit - An analysis of how medical education research(ers) describe how they connect to theory. BMC Med Educ. 2017;17(1):18. DOI: 10.1186/s12909-016-0848-1
42. Nichols TR, Wisner, PM, Cripe G, Gulabchand L. Putting the Kappa Statistic to Use. Qual Assur J. 2010;13:57-61. DOI: 10.1002/qaj.481
43. Souvignier E, Gold A. Wirksamkeit von Lehrmethoden. In: Schweizer K, editor. Leistung und Leistungsdiagnostik. Berlin: Springer; 2006. p.146-166. DOI: 10.1007/3-540-33020-8_10
44. Tenorth HE, Tippelt R, editors. Lexikon Pädagogik. Weinheim: Beltz; 2007. p.137f.
45. Döring N. Forschungsmethoden und Evaluation in Sozial- und Humanwissenschaften. 6. Aufl. Berlin: Springer; 2023. DOI: 10.1007/978-3-662-64762-2

Corresponding author:

PD Dr. Katrin Schüttpelz-Brauns
Heidelberg University, Medical Faculty Mannheim,
Theodor-Kutzer-Ufer 1-3, D-68167 Mannheim, Germany
katrin.schuettpelz-brauns@medma.uni-heidelberg.de

Please cite as

Schüttpelz-Brauns K, Schneider A, Fabry G, Matthes J, Himmelbauer M, Buss B, Giesler M. What is medical education research? An analysis and definition of subjects, objectives and types of research based on articles that have undergone a peer review process. GMS J Med Educ. 2026;43(1):Doc12.
DOI: 10.3205/zma001806, URN: urn:nbn:de:0183-zma0018065

This article is freely available from

<https://doi.org/10.3205/zma001806>

Received: 2024-10-25

Revised: 2025-02-21

Accepted: 2025-04-16

Published: 2026-01-15

Copyright

©2026 Schüttpelz-Brauns et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Was ist Medizinische Ausbildungsforschung? Eine Analyse und Begriffsbestimmung von Gegenständen, Zielen und Typen der Forschung anhand von Artikeln, die einen Peer-Review-Prozess durchlaufen haben

Zusammenfassung

Hintergrund: Medizinische Ausbildungsforschung (MAF) soll einen Beitrag zum Erkenntnisgewinn sowie zur Weiterentwicklung der Ausbildungspraxis leisten. Seit Jahren wird jedoch die mangelnde Relevanz der in diesem Bereich durchgeführten Studien kritisiert. In der vorliegenden Arbeit sollen daher die Aufgaben und Ziele der MAF sowie die Art und Weise der Forschung in diesem Bereich geklärt werden. Hierzu werden die bisher in der MAF üblichen Gegenstände, Ziele und Typen der Forschung analysiert und Kategorien für diese entwickelt.

Methode: Die Entwicklung der Kategorien für *Gegenstände* und *Forschungsziele* erfolgte iterativ in drei Phasen mit mehreren Durchgängen anhand von Stichproben peer-begutachteter Artikel. Je nach Durchgang waren zwei bis sechs Personen in die unabhängige Kategorisierung, Konsentierung und Weiterentwicklung der Kategorien involviert. Zeitgleich wurden *Forschungstypen* für die MAF definiert.

Ergebnisse: 169 Artikel wurden beurteilt. Elf *Gegenstandskategorien* sowie acht Kategorien von Forschungszielen wurden identifiziert und vier als relevant erkannte *Forschungstypen* definiert.

Diskussion: Die gefundenen Kategorien für Gegenstände decken sich z. T. mit vorhandenen Kategoriensystemen, erweitern diese jedoch. Die identifizierten *Forschungsziele* sind spezifischer formuliert als es bisher der Fall war und präzisieren somit den Interpretationsspielraum.

Schlussfolgerung: Die entwickelten Kategoriensysteme können helfen, *Gegenstände* und *Forschungsziele* der Medizinischen Ausbildungsforschung genauer zu definieren sowie die *Forschungstypen* und ihre Bedeutung voneinander abzugrenzen. Zudem sind über die gefundenen Kategorien auch Trends und kurzlebige Phänomene abbildbar.

Schlüsselwörter: Ausbildungsforschung, Forschung, Medizinische Ausbildung, Grundlagenforschung, angewandte Forschung, anwendungsorientierte Grundlagenforschung

Katrin

Schüttpelz-Brauns¹

Achim Schneider²

Götz Fabry³

Jan Matthes⁴

Monika Himmelbauer⁵

Beatrice Buss⁶

Marianne Giesler⁷

1 Medizinische Fakultät
Mannheim der Universität
Heidelberg, Mannheim,
Deutschland

2 Universität Ulm, Medizinische
Fakultät, Ulm, Deutschland

3 Universität Freiburg,
Medizinische Fakultät,
Freiburg i. Brsg., Deutschland

4 Universität zu Köln,
Medizinische Fakultät, Köln,
Deutschland

5 Medizinische Universität
Wien, Wien, Österreich

6 Berner Fachhochschule,
Department Gesundheit,
Bern, Schweiz

7 Freiburg i. Brsg., Deutschland

1. Einführung

Die Ausbildungsforschung an Hochschulen hat ihre Anfänge in den USA der 1920er Jahre [1]. In der Medizin hat sie sich dort ab den 50er Jahren unter der Bezeichnung *medical education research* etabliert [2], [3]. Medical Education Research war eine Reaktion auf verschiedene soziohistorische Faktoren, wie

„...the increasing importance of scientific research, the availability of funds for MER [medical education research], the explosive growth of medical knowledge, and concerns about accountability for, and control of, medical education.“ [4].

Seit Jahren wird in internationalen Zeitschriften immer wieder die mangelnde Relevanz der in der Medizinischen Ausbildungsforschung betriebenen Forschung kritisiert, die zusammenfassend betrachtet einerseits im fehlenden Beitrag zur Wissenschaft im Sinne eines Erkenntnisgewinns, andererseits im Fehlen von Implikationen für die Praxis zum Ausdruck kommt. Die Kritik am fehlenden Erkenntnisgewinn bezieht sich z. B. auf irrelevante Themen oder auf die fehlende Generalisierbarkeit von Ergebnissen, die Kritik der fehlenden praktischen Implikationen z. B. auf die Anwendbarkeit der Ergebnisse auf die Praxis, aber auch auf die gemessenen Outcomes. In Tabelle 1 sind die einzelnen Kritikpunkte dargestellt und erläutert. Um nachvollziehen zu können, wie Forschungsvorhaben der Medizinischen Ausbildungsforschung zum wissenschaftlichen Erkenntnisgewinn bzw. zur Weiterentwicklung der Praxis beitragen könnten, lohnt es sich, die Medizinische Ausbildungsforschung näher zu betrachten. Als erstes stellt sich die Frage nach den konkreten Themen bzw. Gegenständen, die in der Medizinischen Ausbildungsforschung beforscht werden (*Gegenstände*). Auch ist von Relevanz, welche Arten von Forschungszielen verfolgt werden (*Forschungsziele*) und welche Arten von Forschung (*Forschungstyp*) betrieben wird. In den nachfolgenden Abschnitten werden diese drei Bereiche der Medizinischen Ausbildungsforschung näher beleuchtet.

1.1. Gegenstände der Medizinischen Ausbildungsforschung

Schaut man sich einschlägige Lehrbücher an, dann wird die große Bandbreite von *Gegenständen* der Medizinischen Ausbildungsforschung ersichtlich [5], [6], [7], [8]. Diese Vielfalt zeigt sich auch in Artikeln, in denen die *Gegenstände* von Studien der Medizinischen Ausbildungsforschung systematisch analysiert und kategorisiert werden [9], [10]. Nicht immer sind diese Kategorien deckungsgleich. Da zudem einzelne *Gegenstände* aus unterschiedlichen Perspektiven betrachtet bzw. beforscht werden können, ist eine eindimensionale Aufzählung der *Gegenstände* schwer möglich.

Da es unseres Wissens nach keine allgemein anerkannte Systematik von *Gegenständen* in Studien der Medizinischen Ausbildungsforschung gibt, könnte für eine künftige Kategorisierung von *Gegenständen* in der Medizinischen

Ausbildungsforschung ein in der Bildungspsychologie entwickeltes Strukturmodell [11] hilfreich sein. In diesem Modell sind die *Gegenstände* der Bildungspsychologie entlang von drei Dimensionen angeordnet. Diese sind: die *Aufgabenbereiche* Forschung, Beratung, Prävention, Intervention sowie Bildungsmonitoring und Evaluation, die *Bildungskarriere*, die die gesamte Lebensspanne umfasst und somit das lebenslange Lernen hervorhebt, sowie die *Handlungsebenen* (Mikro, Meso, Makro), auf denen die Aufgaben zu leisten sind. Ausgehend von diesem Modell könnten die *Gegenstände* der Medizinischen Ausbildungsforschung wie beispielsweise Studierenden Auswahl, Lernen, Lehren, Prüfung, Curriculum der Dimension *Aufgabenbereiche* zugeordnet werden, die jeweils auf unterschiedlichen Ebenen (Individuum, Gruppen, Organisation) aber auch bezogen auf die *Bildungskarriere* (Ausbildung, Fortbildung, Weiterbildung) wissenschaftlich analysiert werden könnten.

1.2. Forschungsziele der Medizinischen Ausbildungsforschung

In der medizinischen Ausbildung gibt es zwei Rahmenwerke, die *Forschungsziele* der Medizinischen Ausbildungsforschung aufgreifen. Nach Cook et al. zielen Studien auf die Beschreibung (*Description*, Konzentration auf die Beobachtung), die Rechtfertigung (*Justification*, „Wie wirkt eine Maßnahme?“) oder die Erklärung (*Clarification*, „Wie funktioniert etwas?“, „Warum funktioniert etwas?“) ab [12]. In ihrem *Research Compass* fassen Ringsted et al. die *Forschungsziele* unter den Kategorien Modellierung (*Modelling*), Rechtfertigung (*Justifying*), Vorhersage (*Predicting*) und Implementierung (*Implementing*) zusammen [13]. In beiden Beiträgen liegt der Fokus jedoch auf der Art der Datenerhebung bzw. des Studiendesigns, mit denen diese Fragen beantwortet werden sollen. Wie die Kategorien der *Forschungsziele* entstanden sind, wird offengelassen.

1.3. Forschungstypen der Medizinischen Ausbildungsforschung

Betrachtet man die Literatur zur Einordnung von *Forschungstypen* in der Medizinischen Ausbildungsforschung, so gibt es eine Reihe von Forschenden, die diese auf einem eindimensionalen Kontinuum anordnen. So stellte Albert, der die Diskussionen in der Medizinischen Ausbildungsforschung aus soziologischer Sicht unter Zuhilfenahme des Feldkonzepts von Pierre Bourdieu analysierte, fest, dass Medizinische Ausbildungsforschung von zwei Interessengruppen betrieben wird [14]. Die eine Gruppe strebt danach, das Wissen des Forschungsgebietes zu erweitern und zu vertiefen (Erkenntnisinteresse). Diese Interessengruppe repräsentiert die *Grundlagenforschung*. Studien dieses *Forschungstyps* zielen darauf ab, allgemeine Phänomene und ihre Beziehungen grundlegend zu verstehen [13], [15], [16]. Damit soll im Rahmen der Medizinischen Ausbildungsforschung das Wissen und

Tabelle 1: Kritikpunkte von Studien der Medizinischen Ausbildungsforschung

Kritikpunkt	Aspekte	Erläuterungen
Erkenntnisgewinn	Irrelevante Themen bzw. Themen von marginalem Interesse [19], [36], [37], [38]	Themen sind nicht oder nur marginal relevant für die Weiterentwicklung der medizinischen Ausbildung
	Opportunistisch gewählte Studie [17], [20], [39]	Durchführung von Studien, weil sich die Gelegenheit ergibt (z. B. bei Umstellung von Lehrformaten), unabhängig vom aktuellen Wissensstand
	Mehr Beschreibung als Erklärung [9], [12]	Beschreibung von neuen Lehr- oder Assessmentmethoden, statt Untersuchung allgemeiner Wirkprinzipien
	Fehlende Generalisierbarkeit [21], [36], [40]	Generalisierbarkeit der Ergebnisse nicht gegeben
	Fehlende konzeptionelle Einbettung [17], [20], [22], [36], [39], [41]	Kein Bezug zum aktuellen Wissensstand bzw. zu theoretischen Rahmenwerken
Praktische Implikationen	Fehlende praktische Anwendbarkeit von Ergebnissen [36]	Keine Implikation für die Ausbildungspraxis
	Zufriedenheit statt Patienten-Outcome [20], [22]	Wirksamkeitsprüfung nur auf der ersten Stufe des Kirkpatrick-Modells, vierte Stufe gewünscht

das Verständnis von Lernen und Ausbildung vertieft werden [13]. Der Fokus liegt hierbei auf der Frage „Warum funktioniert etwas?“ [13], [16]. Die Ableitung der Forschungsfrage muss daher in das bisherige Wissen zu dem Thema und ggf. auch in einen theoretischen Rahmen eingebettet sein. Die andere Interessengruppe ist bestrebt, mit Hilfe von Forschung praktische Probleme in der medizinischen Ausbildung zu lösen [14] (Anwendungsinteresse). Diese Gruppe repräsentiert die *angewandte Forschung*. Studien dieses *Forschungstyps* zielen darauf ab Probleme für die Praxis zu lösen [17]. Diese können sowohl die Effektivität der Maßnahmen auf das Lernen betreffen [18], als auch die Wirkung von Maßnahmen auf die Gesellschaft bzw. auf die Patient*innen [19], [20], [21], [22]. Bei der Relevanz der Forschung aus Sicht der angewandten Forschenden geht es vor allem um die praktischen Implikationen der Forschungsergebnisse zur Verbesserung der medizinischen Ausbildung [13].

Die eindimensionale Betrachtung mit den Extremen Erkenntnisinteresse vs. Anwendungsinteresse gibt es auch in anderen Forschungsdisziplinen und wurde von Stokes ausführlich beschrieben [23]. Stokes weist jedoch darauf hin, dass die Ziele von Studien zweidimensional betrachtet werden sollten. Diese Dimensionen repräsentieren das Vorhandensein bzw. das Nichtvorhandensein von *Erkenntnisinteresse* und *Anwendungsinteresse*. In seinem zweidimensionalen Quadrantenmodell lassen sich zunächst drei unterschiedliche Forschungstypen unterscheiden, mit denen verschiedene Ziele verfolgt werden [23]. In der *reinen Grundlagenforschung* (*pure basic research*) werden Studien durchgeführt, um Phänomene im wissenschaftlichen Feld zu verstehen. Hier geht es um den reinen Erkenntnisgewinn ohne Anspruch auf praktische Anwendung [23]. Die *angewandte Forschung* (*pure applied research*) zielt auf einen individuellen, Gruppen- oder gesellschaftlichen Bedarf oder die Anwendung ab. Hier geht es um das reine Anwendungsinteresse ohne Anspruch eines Erkenntnisgewinns [23]. Durch die zwei-

mensionale Darstellung entsteht ein dritter Quadrant, bei dem sowohl *Erkenntnisinteresse* als auch *Anwendungsinteresse* hoch sind. Dieser Quadrant repräsentiert die *anwendungsorientierte Grundlagenforschung* (*use-inspired basic research*). Bei diesem Forschungstyp werden unbekannte Grundlagen getestet und gesellschaftliche Bedürfnisse erfüllt [23].

Für drei der Quadranten gibt Stokes konkrete Beispiele an. Der vierte Quadrant ist zwar unbenannt, aber keineswegs leer. Diesem Quadranten ordnet Stokes Forschungsarbeiten zu, die weder von Erkenntnis- noch von Anwendungsinteresse inspiriert sind, sondern stattdessen bestimmte Phänomene systematisch erforschen [23], [24].

1.4. Zielsetzung und Forschungsfragen

Die vorangegangenen Versuche, die Medizinische Ausbildungsforschung anhand ihrer *Gegenstände*, *Forschungstypen* und *Forschungsziele* näher zu beschreiben, um so letztendlich die Bedeutung ihrer Studien bezogen auf ihre Beiträge zur Erkenntnisgewinn und Implikationen für die Praxis aufzuzeigen, hat einige Lücken sichtbar gemacht. So existieren aufgrund der Themen- bzw. Gegenstandsvielfalt keine einheitlichen *Gegenstandskategorien* in der Medizinischen Ausbildungsforschung. Ebenso wenig existieren konkrete Kategorien von *Forschungszielen*, die im Allgemeinen in der Medizinischen Ausbildungsforschung angestrebt werden bzw. angestrebt werden sollten. Und auch wenn die Zuordnung der Studien der Medizinischen Ausbildungsforschung zu den vier möglichen Kategorien von Stokes gut machbar zu sein scheint, liegen keine genauen Kenntnisse vor über die Art der Forschung, die in der Medizinischen Ausbildungsforschung betrieben wird. Um diese Lücken zu füllen, sollen in der vorliegenden Studie die *Gegenstände* und *Ziele* der Medizinischen Ausbildungsforschung sowie die *Art und Weise der Forschung* in diesem Bereich geklärt werden.

Tabelle 2: Überblick über die Phasen und Durchgänge der Studie

DG	Stichprobe	NArtikel	NBeurt	Ziel des Durchgangs (DG)
Phase 1: Induktives Vorgehen				
DG1	SP1	30	2	Anfangskategorien für <i>Forschungsziele</i> und Einordnung von Studien in <i>Forschungstypen</i>
DG2	SP2	50	5	Überprüfung der Kategorien für <i>Gegenstände</i> und <i>Forschungsziele</i> sowie Einordnung der <i>Forschungstypen</i>
Phase 2: Deduktives Vorgehen				
DG3	SP3	30	2	Weiterentwicklung der Kategorien für <i>Gegenstände</i> und <i>Forschungsziele</i> sowie Einordnung der <i>Forschungstypen</i> inkl. Überprüfung der Kategorien
DG4	SP3	12	2	
DG5	SP4	15	2	
DG6	SP5	60	5	Überprüfung der Kategorien für <i>Gegenstände</i> und <i>Forschungsziele</i> sowie Einordnung der <i>Forschungstypen</i> im erweiterten Kreis
DG7	SP1-SP5	50	5	Weiterentwicklung der Kategorien für <i>Gegenstände</i> und <i>Forschungsziele</i> sowie Einordnung der <i>Forschungstypen</i> Überprüfung der Kategorien für <i>Gegenstände</i> und <i>Forschungsziele</i> sowie Einordnung der <i>Forschungstypen</i>
DG8	SP1-SP5	20	6	
Phase 3: Feinjustierung				
DG9	SP1-SP5	20	3	Überarbeitung und Schärfung der Kategorien für <i>Gegenstände</i> , <i>Forschungsziele</i> sowie <i>Forschungstypen</i> mit anschließender Überprüfung
DG10	SP1-SP5	20	3	
DG11	SP1-SP5	20	3	

DG: Durchgang; N_{Artikel}: Anzahl der verwendeten Artikel; N_{Beurt}: Anzahl der beurteilenden Personen zur Überprüfung der Kategorien; SP1-SP5: zufällige Auswahl aus der Grundgesamtheit „alle Artikel aus den bisherigen Durchgängen, welche in einer Zeitschrift aus der medizinischen Ausbildung veröffentlicht wurde unter Ausschluss von Duplikaten“

Hierzu sollen folgende drei Fragen beantwortet werden:

1. Welche Gegenstände werden in der medizinischen Ausbildung beforscht?
2. Welche Forschungsziele werden mit der Medizinischen Ausbildungsforschung verfolgt?
3. Welche im Quadrantenmodell nach Stokes [23] aufgeführten Forschungstypen lassen sich für die Medizinische Ausbildungsforschung ableiten?

2. Methoden

Es ist anzunehmen, dass peer-reviewte Artikel aus der Medizinischen Ausbildungsforschung, die mögliche Bandbreite von *Gegenständen*, *Forschungszielen* und *Forschungstypen* aus Sicht der verfassenden, begutachtenden und herausgebenden Personen wiedergeben. Um diese Annahme zu überprüfen, wurde ein an die Dokumentenanalyse [25] angelehntes Verfahren gewählt. In drei Phasen mit mehreren Durchgängen und Stichproben wurden ausgewählte Artikel den Bereichen *Gegenstand*, *Forschungsziele* und *Forschungstypen* zugeordnet. In Phase 1 (*induktives Vorgehen* [26]) wurden für diese Bereiche Kategorien entwickelt, denen die Studien zugeordnet werden konnten. Phase 2 (*deduktives Vorgehen* [26]) wurde genutzt, um die Kategorien mehrfach zu überarbeiten. In Phase 3 (*Feinjustierung*) wurden die Kategorien und ihre Erläuterungen geschärft. Ein Überblick über alle Phasen gibt Tabelle 2.

2.1. Stichproben

Insgesamt gab es fünf Stichproben (SP1 bis SP5) von Artikeln, die sich bezüglich der Quelle sowie den Einschluss- und Ausschlusskriterien unterscheiden.

Die Stichproben wurden jeweils unabhängig voneinander gezogen, um bei der Weiterentwicklung von Kategorien bzw. deren Überprüfung ein möglichst breites Feld abdecken zu können. Gegen Ende der zweiten Phase stellte sich heraus, dass eine wiederholte Ziehung weiterer Artikel zu keinen neuen Erkenntnissen geführt hätte, sodass die Artikel aus SP1 bis SP5 in einer Datei zusammengefasst wurden. Dabei wurden doppelt vorhandene Artikel anhand der DOI identifiziert und entfernt. Aus der resultierenden Datei wurden im weiteren Verfahren jeweils zufällig Artikel ausgewählt unter Ausschluss der Studien aus rein medizinischen Fachzeitschriften. Eine detaillierte Beschreibung der Stichprobenauswahl befindet sich in Anhang 1.

2.2. Vorgehen

Die Entwicklung der Kategorien erfolgte iterativ (siehe Abbildung 1) über mehrere Durchgänge (DG) hinweg (siehe Tabelle 2). Die Kategorien für *Gegenstände* und *Forschungsziele* wurden dabei inhaltlich erarbeitet, während die Kategorien für die *Forschungstypen* nach Stokes [23] vorgegeben waren, und lediglich für die Medizinische Ausbildung angepasst werden mussten.

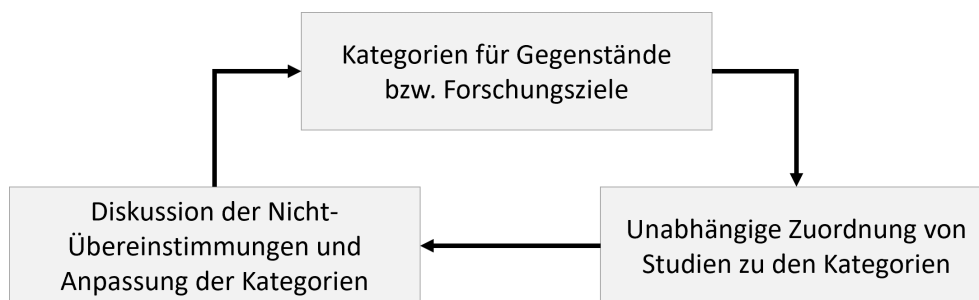


Abbildung 1: Iteratives Vorgehen bei der Entwicklung der Kategorien für Gegenstände und Forschungsziele der Medizinischen Ausbildungsforschung

Tabelle 3: Anfangskategorien für Gegenstände in der Medizinischen Ausbildungsforschung

Dimension	Anfangskategorien
Aufgabenbereiche	Studierendenauswahl, Lernen, Lehren, Prüfung, Curriculum, Begleitung/Führung, Management
Bildungskarriere	Ausbildung, Fortbildung, Weiterbildung
Handlungsebenen	Individuum, Gruppen, Organisation

Vier der beurteilenden Personen kommen aus der Psychologie (MG, MH, AS, KSB), zwei aus der Medizin (GF, JM) und eine aus der Physiotherapie (BB).

2.2.1. Phase 1: Induktives Vorgehen (DG1-DG2)

DG1 zielte darauf ab, Anfangskategorien für *Forschungsziele* zu finden und Artikel den *Forschungstypen* nach Stokes [23] zuzuordnen (siehe Tabelle 2). Hierzu wurden zunächst von zwei Autorinnen (MG, KSB) unabhängig voneinander jeweils das *Forschungsziel* des Artikels, das Erkenntnisinteresse sowie das Anwendungsinteresse paraphrasiert. Aus der Kombination aus Erkenntnis- und Anwendungsinteresse wurde der *Forschungstyp* abgeleitet. Nicht-Übereinstimmungen bei den *Forschungszielen* bzw. adressierte *Forschungstypen* wurden diskutiert, bis eine Einigung in den Einschätzungen erzielt wurde. Das induktive Verfahren wurde gewählt, um unabhängig von den Kategorien von Cook [12] und Ringsted [13] *Forschungsziele* zu entwickeln und ggf. diese mit unterschiedlichen Methoden im Sinne einer Triangulation zu validieren.

Das Ziel für DG2 war die Überprüfung der Kategorien für *Gegenstände*, *Forschungsziele* und *Forschungstypen* durch fünf unabhängige beurteilende Personen (BB, MH, JM, AS, KSB). Die Kategorien für *Gegenstände* und *Forschungsziele* sowie die Einteilung in die *Forschungstypen* waren dabei vorgegeben. Das oben beschriebene Strukturmodell der Bildungspsychologie [11] bildete die Basis der Entwicklung der Anfangskategorien für *Gegenstände* der Medizinischen Ausbildungsforschung. Den einzelnen Dimensionen dieses Modells wurden die in der Medizinischen Ausbildungsforschung gängigen Studiengegenstände zugeordnet (siehe Tabelle 3). Die Kategorien für die *Forschungsziele* ergaben sich aus den Ergebnissen aus DG1. Da davon auszugehen war, dass sich die *Forschungsziele* den Kategorien der *Forschungstypen* zuordnen lassen, war die jeweilige Zuordnung der *Forschungsziele* zu den *Forschungstypen* vorgegeben.

Auf forschungsökonomischen Gründen erfolgte die Zuordnung von Artikeln zu beurteilenden Personen nach einem balancierten Plan (siehe Anhang 2).

Nicht-Übereinstimmungen der Zuordnung zu den Kategorien wurden identifiziert. Falls möglich, wurden die Kategorien unter Berücksichtigung der Anmerkungen zu den einzelnen *Gegenständen* bzw. *Forschungszielen* im Nachgang neu zugeordnet.

2.2.2. Phase 2: Deduktives Vorgehen (DG3-DG8)

DG3-DG5 dienten der Weiterentwicklung der Kategorien für *Gegenstände*, *Forschungsziele* sowie der Einordnung in die *Forschungstypen* (siehe Tabelle 2). Die Kategorien waren vorgegeben und ergaben sich aus den Ergebnissen des jeweiligen vorherigen DG. Zwei Beurteilerinnen (MG, KSB) ordneten unabhängig voneinander die Artikel jeweils den Kategorien zu. Nicht-Übereinstimmungen der Kategorien wurden diskutiert, Kategorien ggf. umbenannt und neu geordnet sowie Erläuterungen zu den einzelnen Kategorien eingefügt.

Ziel von DG6 war es, die in DG5 konsentierten Kategorien inklusive der Erläuterungen für *Gegenstände* und *Forschungsziele* sowie die Einordnung zu den *Forschungstypen* zu überprüfen. Hierfür beurteilten fünf Personen (MG, MH, JM, AS, KSB) unabhängig voneinander die Artikel. Auch hier wurde aus forschungsökonomischen Gründen ein balancierter Plan für die Zuordnung der Beurteilerinnen und Beurteiler zu Artikeln verwendet. Dieser Plan ist äquivalent mit dem aus DG2, nur, dass in jeder Zelle sechs statt fünf Artikel beurteilt wurden.

Sowohl in DG7 als auch in DG8 wurden die Kategorien bei Nicht-Übereinstimmung von MG, MH, JM, AS und KSB besprochen und angepasst. Vor DG 8 analysierten MG und GF jeweils aus psychologischer und medizinischer Sicht, die als kritisch angesehenen Artikel, d.h., solche Artikel, die eine Überschneidung von mehreren *For-*

schungstypen aufwiesen. Der Fokus lag dabei auf den Forschungszielen und ihren Zusammenhang mit den Forschungstypen.

2.2.3. Phase 3: Feinjustierung (DG9-DG11)

In der Phase der Feinjustierung wurden die Kategorien und Erläuterungen für *Gegenstände*, *Forschungsziele* und *Forschungstypen* geschärft (siehe Tabelle 2). Auch hier erfolgten mehrere DG mit Diskussion der Nicht-Übereinstimmungen, Überarbeitung bzw. Schärfung der Kategorien und unabhängige Beurteilung jeweils mit drei Beteiligten (MG, AS, KSB).

2.3. Analyse der Beurteilungsübereinstimmungen

Zur Feststellung der Beurteilungsübereinstimmung wurden Kappa-Maße berechnet, die als zufallsbereinigt anzusehen sind [27], [28]. Zur Ermittlung der Übereinstimmung von zwei beurteilenden Personen wurde Cohens Kappa (κ) berechnet und zur Ermittlung der Übereinstimmung von mehr als 2 Personen Fleiss κ .

In Phasen 1 und 2 wurden für die Einschätzungen der zehn bzw. 18 bewerteten Paare Cohens κ ermittelt. Diese wurden für alle Kategorien des Bereichs *Gegenstände*, der *Forschungsziele* und der *Forschungstypen* berechnet. In einem zweiten Schritt wurde für alle Bereiche jeweils der Median dieser Werte als Schätzung der durchschnittlichen Übereinstimmung zwischen allen beurteilenden Personen bestimmt, da der Median in diesem Fall das Maß ist, das die Gesamtstichprobe am besten charakterisiert [27]. In Phase 1 wird im Ergebnisteil auf die Angabe der Signifikanzen verzichtet, da die ermittelten κ -Werte lediglich jeweils auf den Einschätzungen von fünf Artikeln beruhen.

In Phase 3 wurde Fleiss κ berechnet, da die vorgegebenen Artikel durch drei Personen einzuschätzen waren. Mit diesem Maß kann nicht nur die Beurteilungsübereinstimmung des gesamten Kategoriensystems bestimmt werden, sondern auch für jede einzelne Kategorie.

Als Interpretationshilfe zur Beurteilung des Ausmaßes an Übereinstimmung gelten nach Landis und Koch folgende Angaben: $\kappa < 0,00$ schlechte Übereinstimmung, $0,00-0,20$ geringe Übereinstimmung, $0,21-0,40$ ausreichende Übereinstimmung, $0,41-0,60$ mittelmäßige Übereinstimmung, $0,61-0,80$ beachtliche Übereinstimmung und $0,81-1,00$ fast vollkommene Übereinstimmung [28]. Bei den Beurteilungen mit drei Personen wurden die Beurteilungsrunden so lange durchgeführt bis ausreichende Beurteilungsübereinstimmungen erreicht wurden.

3. Ergebnisse

3.1. Stichprobenbeschreibung

Über alle Stichproben hinweg wurden 197 Artikel ausgewählt, wovon 28 in zwei Stichproben vorhanden waren. Daher wurden insgesamt 169 Artikel beurteilt. Im Laufe des Beurteilungsprozesses wurden zwei Artikel identifiziert, die nicht den Kriterien einer Studie entsprachen: In einem Artikel wurden eine Forschungsmethode und ihre Anwendung in der Medizinischen Ausbildung dargestellt. Bei dem anderen Artikel handelte es sich um einen Brief an den Herausgeber. Beide Artikel wurden aus den weiteren Analysen ausgeschlossen.

Angaben zu den Artikeln aus den einzelnen Stichproben finden sich in Tabelle 4.

3.2. Entwicklung der Kategorien

Da in dieser Studie ein iteratives Vorgehen gewählt wurde, werden die Ergebnisse der einzelnen Phasen an dieser Stelle, nicht wie in wissenschaftlichen Studien üblich, nur beschreibend berichtet, sondern genutzt, um das weitere Vorgehen zu begründen.

Am Ende von Phase 1 wurden für zehn Paare, die Artikel bewertet haben, die folgenden Beurteilungsübereinstimmungen ermittelt:

- *Gegenstandsbereiche*
 - *Kategorie Aufgabenbereiche*: $\kappa = -0,111-0,737$; $\kappa_{\text{Md}} = 0,143$
 - *Kategorie Bildungskarriere*: $\kappa = -0,250-1,000$; $\kappa_{\text{Md}} = 0,302$
 - *Kategorie Handlungsebenen*: $\kappa = -0,071-0,412$; $\kappa_{\text{Md}} = 0,185$
- *Forschungsziele*: $\kappa = -0,087-0,800$; $\kappa_{\text{Md}} = 0,243$
- *Forschungstypen*: $\kappa = -0,364-0,444$; $\kappa_{\text{Md}} = 0,000$.

Während der Phase des induktiven Vorgehens (Phase 1) zeigte sich, dass sich teilweise die den Dimensionen des Strukturmodells der Bildungsforschung (*Aufgabenbereiche*, *Bildungskarriere*, *Handlungsebenen*) zugeordneten *Gegenstände* überschneiden. So waren z. B. die *Gegenstände* Lehren, Lernen und Curriculum nicht immer trennscharf voneinander zu unterscheiden. Zudem konnten einige *Gegenstände* den Dimensionen *Handlungsebene* und *Bildungskarriere* nicht eindeutig zugeordnet werden. Ausgehend von diesen Überschneidungen, die zum Teil die geringen Beurteilungsübereinstimmungen verursacht haben könnten, wurde das Konzept von drei Dimensionen aufgegeben und alle *Gegenstände* zu einem Gegenstandsbereich zusammengefasst. Sowohl für die *Gegenstände* als auch für die *Forschungsziele* wurde in den Durchgängen der Phase 2 die Kategorien sukzessive im Konsensverfahren angepasst. Dies betraf die Kategorien selbst als auch die Erläuterungen zu diesen. Als besonders schwierig erwies sich die Konsensfindung, da die Beurteilerinnen und Beurteiler dieser Studie wegen ihrer unterschiedlichen fachlichen Sozialisation verschie-

Tabelle 4: Beschreibung der einzelnen Stichproben

Stichprobe	Gezogen am	Zeitraum der Veröffentlichungen	Herkunft der Artikel N (%)	Übersichtsarbeiten N (%)
SP1	02.12.2021	2021 ¹	Med Educ J = 30 ² (100%)	4 (13%)
SP2	15.12.2021	2011-2021 Q ₂₅ = 2016 Q ₅₀ = 2018 Q ₇₅ = 2019	Med Educ J = 37 (74%) Med J = 13 (26%)	16 (32%)
SP3	15.12.2021	2014-2019 Q ₂₅ = 2014 Q ₅₀ = 2017 Q ₇₅ = 2019	Med Educ J = 41 (98%) Med J = 1 (2%) ³	8 (19%)
SP4	30.10.2022	2005-2022 Q ₂₅ = 2012 Q ₅₀ = 2018 Q ₇₅ = 2020	Med Educ J = 15 (100%)	5 (33%)
SP5	13.02.2023	2003-2022 Q ₂₅ = 2014 Q ₅₀ = 2017 Q ₇₅ = 2021	Med Educ J = 60 (100%)	6 (10%)

Q₂₅: 25%-Perzentil, Q₅₀=Median, Q₇₅=75%-Perzentil; Med Educ J: Zeitschrift aus der Medizinischen Ausbildung, Med J: medizinische Zeitschrift;

¹ davon 7 vorab online verfügbar und erst im Jahr 2022 endgültig veröffentlicht;

² davon 17 Artikel (57%) aus dem deutschsprachigen Raum;

³ ist erst nachträglich identifiziert worden.

dene Vorstellungen über Wissenschaft hatten. Daneben gab es Unschärfen bei den Begriffen sowie fehlende und sich überschneidende Kategorien. Zudem zeigte sich, dass die Dimensionen Erkenntnis- und Anwendungsinteresse im Quadrantenmodell nach Stokes [23] unterschiedlich interpretiert werden können. Unter Zuhilfenahme von zusätzlicher Literatur [23], [29], [30], [31], [32] und der Diskussion konkreter Studien wurden die Forschungstypen stärker voneinander abgegrenzt.

Für den letzten Durchgang von Phase 2 wurden für 18 Paare, die Artikel bewertet haben, folgende Cohens Kappa berechnet:

- Gegenstände: $\kappa = -0,053 - 0,722$; $\kappa_{\text{Md}} = 0,214$
- Forschungsziel: $\kappa = -0,111 - 0,643$; $\kappa_{\text{Md}} = 0,268$
- Forschungstyp: $\kappa = -0,176 - 1,000$; $\kappa_{\text{Md}} = 0,354$

Im Zuge der Feinjustierung wurde der vierte Quadrant wie von Bush [33] vorgeschlagen als *background research* (Hintergrundforschung) umbenannt. Diese Änderung ging mit einer höheren Beurteilungsübereinstimmung einher. Am Ende dieser Phase waren die Beurteilungsübereinstimmungen wie folgt: für Gegenstände $\kappa_{\text{Fleiss}} = 0,63$, $p < 0,001$, für Forschungsziel $\kappa_{\text{Fleiss}} = 0,52$, $p < 0,001$ und für Forschungstyp $\kappa_{\text{Fleiss}} = 0,58$, $p < 0,001$. Die Ergebnisse zu den einzelnen Kategorien sind in Anhang 3, Tabellen A1-A3 aufgeführt.

3.3. Gegenstände, Forschungsziele und Forschungstypen der Medizinischen Ausbildungsforschung

Nach Abschluss der drei Phasen stehen nunmehr elf *Gegenstandskategorien*, acht Kategorien von *Forschungs-*

zielen sowie Erläuterungen zu den *Forschungstypen* zur Verfügung (siehe Tabelle 5, Tabelle 6 und Tabelle 7).

4. Diskussion

Ausgangspunkt für diese Studie ist die oftmals beklagte fehlende Relevanz der Medizinischen Ausbildungsforschung. Denn der Medizinischen Ausbildungsforschung wird häufig vorgeworfen mit ihren Studien kaum zum Erkenntnisgewinn in diesem Bereich beizutragen oder aber auch Erkenntnisse zu produzieren, die wenig praxisbezogen sind.

Um besser nachvollziehen zu können, wie Forschungsvorhaben in der Medizinischen Ausbildungsforschung Beiträge zur Wissenschaft im Sinne von Erkenntnisgewinn oder zur Weiterentwicklung der Praxis leisten können, wurden in dieser Studie folgende Fragen gestellt:

1. Welche Gegenstände werden in der medizinischen Ausbildung beforscht?
2. Welche Forschungsziele werden mit der Medizinischen Ausbildungsforschung verfolgt?
3. Welche im Quadrantenmodell nach Stokes [23] aufgeführten Forschungstypen lassen sich für die Medizinische Ausbildungsforschung ableiten?

Um diese Fragen zu beantworten, wurden Forschungsartikel, die Peer-Reviews durchlaufen haben, herangezogen und entsprechend analysiert. Dies geschah in einem iterativen Prozess, in dem Ausgangskategorien der einzelnen Bereiche überprüft und angepasst wurden. Dieser Prozess der unabhängigen Überprüfung der Kategorien und An-

Tabelle 5: Gegenstände der Medizinischen Ausbildungsforschung inkl. Erläuterungen

Gegenstand	Erläuterungen
Eigenschaften, Motivation und Verhalten von Lernenden und Lehrenden	Eigenschaften, Bedürfnisse, Motive, Interessen, (Lern-)Leistungen, Haltungen und Verhalten von Studierenden/Lehrenden einschließlich Genderunterschiede und wie Lernende Lehr-/Lernangebote nutzen. Der Schwerpunkt liegt hier auf dem Lernprozess. <i>Beispiele: klinische Entscheidungsfindung, wissenschaftliches Denken, Resilienz, emotionale Intelligenz, professionelle Identitätsentwicklung, Reflexion</i>
Theorien/ Modelle, Konstrukte	Beschreibung bzw. Erklärung der Realität mit Hilfe von Theorien, Modellen und/oder Konstrukten. Eine Theorie ist ein System wissenschaftlich begründeter Aussagen zur Erklärung von Erscheinungen, (natürlichen) Prozessen und der ihnen zugrundeliegenden Gesetzmäßigkeiten (<i>Beispiele: Selbstbestimmungstheorie nach Ryan und Deci, Sozialkognitive Lerntheorie nach Bandura</i>). Ein Modell stellt ein vereinfachtes Abbild der Realität dar oder bietet prinzipielle Herangehensweisen zu einer Problemlösung (<i>Beispiele: Miller-Pyramide, Kern-Zyklus</i>). Ein Konstrukt beschreibt Merkmale oder Zustände, die nicht direkt beobachtbar sind (<i>Beispiele: Intelligenz, Motivation</i>).
Lehrmethode	Didaktische Handlungen und Maßnahmen, die absichtlich, planvoll und bewusst eingesetzt werden, um Lernvorgänge auszulösen oder zu beeinflussen [43]. <i>Beispiele: Unterricht am Krankenbett, Simulationsunterricht, individualisiertes Feedback, Team-based Learning, Peer-assisted Learning</i>
Curriculum; medizinische Ausbildung (auf Meta-Ebene)	Ein Curriculum orientiert sich an Lehrzielen und am Ablauf des Lehr- und Lernprozesses oder des Studiengangs. Insbesondere enthält es auch Aussagen über die Rahmenbedingungen des Lernens [44]. Die vorliegende Definition umfasst auch die nicht offiziell vorgesehene implizite Vermittlung von Lerninhalten und -formen sowie Wirkungen auf die Sozialisation jenseits von Lehrplänen (hidden curriculum). Ein Curriculum kann einen gesamten Studiengang betreffen, aber auch zusammenhängende Lehrangebote, wie Wahlfächer. Zu dieser Kategorie gehört auch die Betrachtung der medizinischen Ausbildung auf der Meta-Ebene, wie das Ausbildungssystem eines Landes. <i>Beispiele: Masterprogramm, Modellstudiengang, Wissenschaftscurriculum, Promotionsprogramm, Hidden Curriculum</i>
Sonstige Interventionen	Interventionen, mit denen körperliches oder seelisches Wohlbefinden unterstützt bzw. gefördert werden soll; Maßnahmen, die zu einer Qualitätsverbesserung führen sollen. <i>Beispiele: Achtsamkeitstrainings</i>
Messinstrumente (inkl. summative Prüfungen), formative Prüfung als Messinstrument	Mittel, um beobachtete Daten in Messwerte zu überführen; Dies betrifft v.a. Messinstrumente für die Studierendenauswahl (z. B. HAM-Nat), zur Erfassung der Leistung (Prüfungen, z. B. OSCE, Script Concordance Test) oder zum Einsatz in der Forschung (z. B. Approaches to Learning and Studying Inventory, ALSI).
Datenanalyseverfahren	Statistische Verfahren zur Generation von Kenngrößen, die quantitativ gewonnen wurden (z.B. Bayes-Statistik, Cumulative Deviation Method) bzw. Verfahren zur Generation zusammenhängender Informationen, die qualitativ gewonnen wurden (z.B. Thematische Inhaltsanalyse nach Braun & Clark)
Qualitätskriterien und -indikatoren	Kriterien und Indikatoren, mit denen die Qualität von Lehrformaten sowie Maßnahmen in der Ausbildung sichergestellt bzw. überprüft werden können <i>Beispiel: Akkreditierungsstandards, Qualitätskriterien für Lehre, Qualitätskriterien für Studien der Medizinischen Ausbildungsforschung</i>
Positionen, Funktionen, Rollen, Karriere	Studien, die sich mit beruflichen Themen befassen, wie beispielsweise <i>Frauen in der Medizin, Karrierewege von Pflegekräften, Führungskompetenzen</i> . Position bezeichnet den formalen Platz (Rechte und Pflichten), den jemand in einem System, in einer Organisation einnimmt (z.B. Fachärztin). Funktion beschreibt den Zweck und die inhaltlichen Aufgaben (Verantwortungsbereich = Aufgaben und Verantwortlichkeit), die mit der Position verbunden sind (z.B. Fachärztin für Chirurgie). Rolle(n) sind das personale Gegenüber der Funktion. Sie beschreiben die Verhaltenserwartungen, die an Funktionsinhaber z.B. Fachärztin für Chirurgie gerichtet werden.
Medizinische Ausbildungsforschung	Studien in diesem Bereich zielen darauf ab Gegenstände der Medizinischen Ausbildungsforschung zu erfassen, die Qualität der Forschungsarbeiten zu ermitteln und ihre Methoden kennen zu lernen, um diese ggfs. zu verbessern bzw. zu erweitern. <i>Beispiele: Status quo der Medizinischen Ausbildungsforschung in Europa, Überblick über die Methoden der Medizinischen Ausbildungsforschung</i>
Lernumgebungen, Lernbedingungen, Lernkontexte	Strukturelle und personelle Kontextfaktoren für das Lernen, wie <i>digitale Lernumgebungen, Architektur der Gebäude, Auslandseinsatz, Ansprechbarkeit und Verfügbarkeit von Lehrpersonen, Usability von digitalen Tools</i>

Tabelle 6: Forschungsziele der Medizinischen Ausbildungsforschung inkl. Erläuterungen und Beispielen

Ziel der Forschung	Erläuterungen
Untersuchung von Eigenschaften, Motivation, Einstellungen und Verhalten (und deren Veränderung) von Studierenden und Lehrenden bzw. eine Mehrwert generierende Zusammenfassung von Erkenntnissen in diesem Bereich	Wie reagieren Studierende/Lehrende auf ihre Umgebung, auf Veränderungen in ihrer Umgebung, auf Anforderungen etc. bzw. wie nehmen sie diese Aspekte wahr? Wie verhalten sich Studierende/Lehrende als Gruppenmitglieder? Wie lernen Studierende? Wie entwickeln sich Merkmale (ohne Interventionen)?
(Weiter-) Entwicklung und / oder empirische Prüfung einer Theorie, eines Modells, eines Konstrukts bzw. eine Mehrwert generierende Zusammenfassung von Erkenntnissen in diesem Bereich	Wie können beobachtete Zusammenhänge theoretisch erklärt werden? Können theoretisch postulierte Zusammenhänge empirisch nachgewiesen werden? Muss ein Modell überarbeitet werden? Wie lassen sich nicht direkt beobachtbare Merkmale und Zustände angemessen beschreiben (Konstrukt)? Wie verstehen Stakeholder ein Konstrukt?
Empirische Überprüfung von Lehrmethoden, Lehrmaterialien bzw. sonstige Maßnahmen bezogen auf Lernleistungen, Motivation, Zufriedenheit, Einstellungen etc. bzw. eine Mehrwert generierende Zusammenfassung von Erkenntnissen in diesem Bereich	Ist ein Lehrmaterial, ein Lehrformat bzw. eine Maßnahme wirksam im Hinblick auf definierte Outcomes? Wie nehmen Studierende Lehrformate etc. wahr?
Entwicklung und Überprüfung von Messinstrumenten für die Studierendenauswahl, zur Erfassung der Leistung (Prüfung), zur Überprüfung wissenschaftlicher Fragestellungen etc. bzw. eine Mehrwert generierende Zusammenfassung von Erkenntnissen in diesem Bereich	Welches Konstrukt soll gemessen werden? Wie wird das Konstrukt operationalisiert? Wie zuverlässig misst das Instrument das, was es messen soll (Reliabilität)? Und misst das Instrument überhaupt das Konstrukt, das es messen soll (Validität)? Wie gut sind weitere Testgütekriterien, wie Testfairness, Praktikabilität etc.? In welchen Bereichen werden welche Fragebögen eingesetzt? Gibt es Unterschiede bezogen auf die Zielgruppe(n)?
(Weiter-)Entwicklung und / oder Überprüfung von Forschungsansätzen, Forschungsmethodik und Datenanalyseverfahren bzw. eine Mehrwert generierende Zusammenfassung von Erkenntnissen in diesem Bereich	Welche Forschungsmethoden sind adäquat, um Fragen der Medizinischen Ausbildungsforschung zu beantworten? Wie zuverlässig und valide sind die Ergebnisse aus einem neu entwickelten bzw. weiterentwickelten Datenanalyseverfahren?
Untersuchung von Bedingungen für den Einsatz von Lehrmethoden, Lehrmaterialien bzw. Maßnahmen (u.a. im Rahmen von Problem- und Machbarkeitsanalysen)	Unter welchen Bedingungen sind gegebene Lehrmethoden, Lehrmaterialien bzw. andere curriculare Maßnahmen wirksam bzw. nicht wirksam? Die Bedingungen können u.a. wie folgt untersucht werden: (1) Umsetzungsarten von Lehrformaten inkl. förderliche- und hinderliche Bedingungen. (2) Ursachen von Problemen untersuchen (Probleme sind bereits bekannt im Gegensatz zur Kategorie Status quo, s.u.). (3) Überprüfung der technisch/wirtschaftlichen Umsetzbarkeit von Maßnahmen etc., z. B. Feasibility-Studien.
Entwicklung von konsensbasierten, normativen Vorgaben und Empfehlungen	Welche Lernziele, Lehrinhalte etc. sollen in einem Curriculum vermittelt werden? Welche Kompetenzen sollten Lehrende haben?
Darstellung eines Status quo	Welche Informationen können über das systematische Sammeln von Daten zu einem Thema gewonnen werden? Ziel ist hierbei einen Überblick über ein Thema zu erhalten, z. B. über die Umsetzung politischer Vorgaben oder um im Ergebnis potenzielle Forschungslücken oder praktische Probleme identifizieren zu können, auch wenn es noch nicht klar ist, ob es überhaupt welche gibt.

passung wechselten einander ab, bis eine zufriedenstellende Beurteilungsübereinstimmung erreicht wurde. Die gefundenen Kategorien zu den *Gegenständen* decken sich zum Teil mit bisherigen Kategorien. So verwenden Raes et al. [9] ähnliche Kategorien mit einer etwas anderen Sortierung, z. B. sind die Lehr-Lernmethoden von Raes et al. bei uns aufgeteilt in Lehrmethoden und Eigenschaften, Motivation und Verhalten von Lernenden und Lehrenden. Die sieben Forschungsthemen von Raes et al. konnten wir bis auf zwei replizieren. Der *Gegenstand*

Fakultätsentwicklung kann in der vorliegenden Kategorisierung je nach Thema den Kategorien Curriculum, Qualitätskriterien und -indikatoren oder Positionen, Funktionen, Rollen, Karriere zugeordnet werden. Entsprechend kann der *Gegenstand* soziale Kompetenz der Kategorie Eigenschaften, Motivation und Verhalten von Lernenden und Lehrenden zugeordnet werden. Verglichen mit den von Rotgans beschriebenen *Gegenstandskategorien* [10] sind keine Übereinstimmungen festzustellen, da die von uns entwickelten Kategorien übergeordnet formuliert

Tabelle 7: Mögliche Forschungstypen der Medizinischen Ausbildungsforschung inkl. Erläuterungen

Forschungstyp	Erläuterungen
Reine Grundlagenforschung (curiosity-inspired pure basic research for understanding)	Das alleinige Ziel dieser Forschung ist ein wissenschaftlicher Erkenntnisgewinn für das Fachgebiet bezogen auf das Verstehen allgemeiner Phänomene, Zusammenhänge oder sogenannter Gesetzmäßigkeiten. <i>Reine Grundlagenforschung</i> führt zu Erkenntnissen, für die es derzeit (noch) keine direkte Anwendung gibt (vgl. [33] S. 89). Im Zentrum der <i>reinen Grundlagenforschung</i> stehen Theorien (vgl. [45], S. 90).
Anwendungsorientierte Grundlagenforschung (use-inspired research for understanding and use)	Forschung mit dem ein praktisches Problem gelöst bzw. praktische Anwendungen untersucht werden sollen aber gleichzeitig wissenschaftlicher Erkenntnisgewinn für das Fachgebiet angestrebt wird. Das bedeutet, dass die Erkenntnisse aus der Studie für die Weiterentwicklung von Theorien verwendet werden können (vgl. [29]).
Angewandte Forschung (use-inspired pure applied research for use)	Forschung, die darauf abzielt, Lösungen für praktische Probleme zu finden, Technologien usw. zu verbessern. Der Fokus liegt hier auf praxisrelevante Erkenntnisse, die stärker auf konkrete Praxisprobleme zugeschnitten und weniger generalisierbar sind (vgl. [45], S. 187). Angewandte Ziele können sein: die Überprüfung der Wirkung/des Nutzens/der Effektivität einer Maßnahme oder die teststatistische Überprüfung eines Instruments.
Hintergrundforschung (background research)	Wissenschaftliches Erheben und Auswerten von Daten, um einen Überblick über ein Thema oder Teilaspekte eines Themas zu erhalten, z. B. Durchführung von Bestandsaufnahmen (vgl. [23], S. 74).

sind. So können der Rubrik Messinstrumente die von Rotgans aufgeführten *Gegenstände* „Issues in student assessment, objective structured clinical examination, clinical competence assessment“ zugeordnet werden. Das in der vorliegenden Arbeit für *Gegenstände* entwickelte Kategoriensystem fügt sich somit gut in die bisherigen Kategorisierungsversuche ein, erweitert diese jedoch und hat den Vorteil, dass auch neue *Gegenstände* in diese Kategorisierung aufgenommen werden können.

Was die *Forschungsziele* der Medizinischen Ausbildungsforschung betrifft, lassen bisherige Klassifikationen Interpretationsspielraum. Sowohl bei Cook [12] als auch bei Ringsted [13] werden zwar Kategorien für *Forschungsziele* aufgelistet. Es wird jedoch weder dargestellt, wie man auf diese *Forschungsziele* kommt, noch, was diese konkret bedeuten. So können einige der von uns untersuchten Studien der Kategorie *Beschreibung* von Cook [12] zugeordnet werden. Beispielsweise gibt es Studien, in denen Phänomene beschrieben werden, die mit bisherigen Theorien nicht zu erklären sind. Diese Studien liefern damit einen Ausgangspunkt für die Weiterentwicklung und empirische Prüfung von Theorien. Andere Studien wiederum, in denen beispielsweise die Verbreitung bestimmter Lehr- oder Prüfungsformate ohne theoretische oder praktische Implikationen dargestellt werden, können gemäß der Kategorisierung von Cook et al. ebenfalls der Kategorie *Beschreibung* zugeordnet werden. In der vorliegenden Kategorisierung für *Forschungsziele* werden jedoch solche Studien der Kategorie *Darstellung eines Status quo* zugeordnet.

Zusammenfassend betrachtet lassen sich die von uns identifizierten *Forschungsziele*, zwar grob in die bisherigen Klassifikationen einordnen. Sie sind jedoch so spezifisch formuliert, dass sie eher eine eindeutige Zuordnung garantieren, da der Interpretationsspielraum weitestgehend

eingeschränkt wurde. Zudem wurden weitere *Forschungsziele* identifiziert, die sich auf die (Weiter-) Entwicklung und Überprüfung von Messinstrumenten und Datenanalyseverfahren sowie auf die Entwicklung von konsensbasierten, normativen Vorgaben und Empfehlungen beziehen.

Bezogen auf die für die Medizinische Ausbildungsforschung als relevant erachteten *Forschungstypen* ließen sich in unserer Untersuchung drei *Forschungstypen* ableiten: *anwendungsorientierte Grundlagenforschung*, *angewandte Forschung* und *Hintergrundforschung*. Eine Kategorie *Grundlagenforschung* konnte dagegen nicht identifiziert werden. Bezugnehmend auf die Forderungen von Eva und Ringsted scheint dieses Ergebnis partiell stimmig zu sein, denn gemäß ihren Ausführungen sollte sich die Medizinische Ausbildungsforschung auf die *anwendungsorientierte Grundlagenforschung* beziehen, also sowohl das *Erkenntnis*- als auch das *Anwendungsinteresse* adressieren. So schreibt Ringsted „Research in medical education seeks to deepen the knowledge and understanding of learning, teaching and education. It is neither about solving concrete, local problems nor about providing general, universal solutions“ ([13], p.695). Eva geht es vor allem um das Verstehen von Phänomenen, die praktisch relevant sind („identify phenomena of interest (and practical relevance)“... „imperative of understanding“, [16], p.295). Um diese praktisch relevanten Phänomene zu verstehen, müssen wir die Prozesse des Lernens, Lehrens und der Bildung untersuchen. Da diese Prozesse nicht direkt beobachtbar sind, treten prüfbare Theorien in den Vordergrund, die diese Prozesse erklären können. Diese Theorien sind jedoch Gegenstand der *reinen Grundlagenforschung* [34]. In der Medizinischen Ausbildungsforschung wird *reine Grundlagenforschung* jedoch noch viel zu selten betrieben. Tolsgaard et al.

ordneten in ihrem systematischen Review lediglich 2% der randomisierten Studien dem Bohr-Quadranten (=reine Grundlagenforschung) zu [35]. Die Seltenheit des Vorkommens solcher Studien mag daher ein Grund sein, warum in der vorliegenden Untersuchung keine Studien gefunden wurden, die der *reinen Grundlagenforschung* zuzuordnen sind. Aufgrund der bereits oben beschriebenen Bedeutung dieser Kategorie, wurden die *Forschungstypen* um die Kategorie *Grundlagenforschung* ergänzt.

4.1. Stärken und Schwächen der Studie

Die Stärke unserer Studie liegt darin, dass bisherige Klassifikationen für *Forschungsziele* einen starken Fokus auf die Forschungsmethodik haben. Die Auswahl der Forschungsmethodik sollte jedoch erst erfolgen, wenn eine relevante Forschungsfrage identifiziert ist. Bei unseren Diskussionen und Zuordnungen spielte die Forschungsmethodik jedoch keine Rolle, sodass wir unabhängig von der Wahl der „richtigen“ Methode die *Forschungsziele* identifizieren konnten. Hervorzuheben ist, dass peer-begutachtete Artikel, von denen anzunehmen war, dass sie relevante Forschung repräsentieren, einem an eine Dokumentenanalyse angelehnten Verfahren unterzogen wurden. Zur Entwicklung der Kategorien und ihren Erläuterungen wurde iterativ vorgegangen, indem sich Phasen der Konsensfindung und der Prüfung der Beurteilungsübereinstimmung abwechselten. Zudem wurden Personen aus unterschiedlichen Wissenskulturen einbezogen.

Durch den iterativen Prozess der Stichprobenziehung, Kategorisierung, Konsentierung, Übereinstimmungskontrolle, Überarbeitung der Kategorisierung an weiteren Artikeln etc. konnten Kategorien entwickelt werden, die auch Trends und zeitlich begrenzten Themen standhalten. So werden Themen wie Künstliche Intelligenz oder COVID-Pandemie zwar nicht wörtlich in den Kategorien aufgeführt, sind jedoch dennoch implizit vorhanden, z. B. als Lehr- oder Prüfungsformat bzw. Veränderung von Studierendenmotivation.

Die Studie weist auch einige Schwächen auf. Vor allem in der Phase der *Feinjustierung* (Phase 3) erschwerten einige schlecht konzipierte Studien die Kategorienbildung, weil z.B. *Gegenstände* und *Forschungsziele* nicht explizit benannt oder in verschiedenen Teilen der Artikel (z.B. Abstract vs. Volltext) nicht stringent adressiert wurden. Diese Inkonsistenzen wirkten sich vermutlich negativ auf die Beurteilungsübereinstimmung aus. Ein weiteres Problem stellte die unterschiedliche Verwendung der Begriffe *Theorie* und *Modell* in manchen Artikeln dar. Zum einen wurden diese Begriffe zur Beschreibung natürlicher Prozesse herangezogen, um diese dann anhand empirischer Daten zu prüfen (epistemische Bedeutung). Zum anderen wurden die Begriffe *Theorie* und *Modell* verwendet, ohne jedoch den Anspruch zu verfolgen, Verhalten und Erleben zu erklären bzw. zu beschreiben. Stattdessen war das Ziel dieser Studien, sich auf einheitliche Vorgaben zu einigen (normative Bedeutung). Auch die Zuordnung zu den *Forschungstypen* wurde manchmal durch Mängel bzw.

Uneindeutigkeiten in den Artikeln beeinträchtigt. Es wurden mitunter *Forschungsziele* beschrieben, die eigentlich der *anwendungsorientierten Grundlagenforschung* hätten zugeordnet werden können, aber wegen des fehlenden Theoriebezugs nicht zugeordnet werden konnten. Zusammenfassend betrachtet, können solche Mängel zu einer Einschränkung des Ausmaßes der Beurteilungsübereinstimmung führen.

Eine weitere mögliche Schwäche der vorliegenden Studie ist, dass einige Kategorien seltener zur Kategorisierung der Artikel gewählt wurden. Dies könnte daran liegen, dass Studien, die sich solchen Kategorien zuordnen lassen, tatsächlich seltener durchgeführt werden. Nicht auszuschließen ist zudem, dass unsere Erläuterungen dieser Kategorien noch zu ungenau sind. Unser Vorgehen war auf die Entwicklung der Kategorien und ihren Erläuterungen fokussiert und nicht darauf, das tatsächliche Vorkommen der Studien in den einzelnen Kategorien festzustellen und zu berücksichtigen.

5. Fazit

Unsere Kategoriensysteme können helfen, *Gegenstände* und *Forschungsziele* der Medizinischen Ausbildungsforschung genauer zu definieren sowie die *Forschungstypen* und ihre Bedeutung besser voneinander abzugrenzen. Damit sollte es möglich sein, relevante Forschungsfragen zu formulieren und ein Vorgehen zu wählen, in dem erklärende Theorien eine Rolle spielen bzw. das (allgemeine) Problem der Praxis im Vordergrund steht, je nachdem, um welchen Forschungstyp es sich handelt. Erst im nächsten Schritt sollte die Forschungsmethodik so gewählt werden, dass mit Hilfe der erhobenen und ausgewerteten Daten die Forschungsfrage beantwortet werden kann. Zusätzlich sollten die verwendeten Begriffe und vor allen deren Bedeutung diskutiert und konsentiert werden, damit zukünftig die Beteiligten das Gleiche meinen und ihre Studien besser zum Erkenntnisgewinn in der Medizinischen Ausbildung und schließlich zur Weiterentwicklung der Praxis beitragen können.

Danksagung

Wir danken cand. med. Olivia Steiger für die Zuarbeiten bei der Literatur.

ORCIDs der Autor*innen

- Katrin Schüttpelz-Brauns: [0000-0001-9004-0724]
- Achim Schneider: [0000-0002-8602-8535]
- Götz Fabry: [0000-0002-5393-606X]
- Jan Matthes: [0000-0003-2754-1555]
- Monika Himmelbauer: [0000-0001-5516-1993]
- Beatrice Buss: [0000-0001-7270-382X]
- Marianne Giesler: [0000-0001-9384-2343]

Interessenkonflikt

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

Anhänge

Verfügbar unter <https://doi.org/10.3205/zma001806>

1. Anhang_1.pdf (137 KB)
Beschreibung zur Stichprobenauswahl
2. Anhang_2.pdf (151 KB)
Balancierter Plan von 50 Artikeln, die in Phase 1 im zweiten Durchgang zur Überprüfung der gefundenen Kategorien dienten
3. Anhang_3.pdf (156 KB)
Beurteilungsübereinstimmungen zum Abschluss der Studie

Literatur

1. McKeachie WJ. Research on College Teaching: The Historical Background. *J Educ Psychol.* 1990;82(2):189-200. DOI: 10.1037/0022-0663.82.2.189
2. Norman G. Fifty years of medical education research: waves of migration. *Med Educ.* 2011;45(8):785-791. DOI: 10.1111/j.1365-2923.2010.03921.x
3. Schmidt HH, Mamede S. How cognitive psychology changed the face of medical education research. *Adv Health Sci Educ Theory Pract.* 2020;25(5):1025-1043. DOI: 10.1007/s10459-020-10011-0
4. Kuper A, Albert M, Hodges BD. The origins of the field of medical education research. *Acad Med.* 2010;85(8):1347-1353. DOI: 10.1097/ACM.0b013e3181dce9a7
5. Swanwick T, Forrest K, O'Brien BC, editors. *Understanding Medical Education: Evidence, Theory, and Practice*, 3rd ed. Hoboken (NJ): Wiley Blackwell; 2018. DOI: 10.1002/9781119373780
6. Dent J, Harden R, Hunt D, editors. *A practical guide for medical teachers*. 6th ed. London, UK: Elsevier; 2021.
7. Norman GR, Vleuten CP, Newble DI, Dolmans DH, Mann KV, Rothman A, Curry L, editors. *International Handbook of Research in Medical Education*. Dordrecht: Springer; 2012.
8. Fabry G. *Medizindidaktik. Für eine kompetenzorientierte, praxisrelevante und wissenschaftlich fundierte Ausbildung*. 2. vollst. überarb. Aufl. Bern: Hogrefe; 2023. DOI: 10.1024/85852-000
9. Raes P, Bauer D, Schöppe F, Fischer MR. The active participation of German-speaking countries in conferences of the Association for Medical Education in Europe (AMEE) between 2005 and 2013: A reflection of the development of medical education research? *GMS Z Med Ausbild.* 2014;31(3):Doc28. DOI: 10.3205/zma000920
10. Rotgans JI. The themes, institutions, and people of medical education research 1988-2010. *Adv Health Sci Educ Theory Pract.* 2012;17(4):515-527. DOI: 10.1007/s10459-011-9328-x
11. Spiel C, Wagner P, Götz T, Lüftenegger M, Schober B. *Bildungspsychologie: Konzeption, Strukturmodell, Stellenwert und Perspektiven*. In: Spiel C, Götz T, Wagner P, Lüftenegger M, Schober B, editors. *Bildungspsychologie – ein Lehrbuch*. 2. vollst. überarb. Aufl. Göttingen: Hogrefe; 2022. p.11-30. DOI: 10.1026/03108-000
12. Cook DA, Bordage G, Schmidt HG. Description, justification and clarification: a framework for classifying the purposes of research in medical education. *Med Educ.* 2008;42(2):128-133. DOI: 10.1111/j.1365-2923.2007.02974.x
13. Ringsted C, Hodges B, Scherpbier A. 'The research compass': an introduction to research in medical education: AMEE Guide no. 56. *Med Teach.* 2011;33(9):695-709. DOI: 10.3109/0142159X.2011.595436
14. Albert M. Understanding the debate on medical education research: a sociological perspective. *Acad Med.* 2004;79(10):948-954. DOI: 10.1097/00001888-200410000-00009
15. Eva K. Broadening the debate about quality in medical education research. *Med Educ.* 2009;43(4):294-296. DOI: 10.1111/j.1365-2923.2009.03342.x
16. Eva K. Publishing during COVID-19: Lessons for health education research. *Med Educ.* 2021;55(3):278-280. DOI: 10.1111/medu.14450
17. Albert M, Hodges B, Regehr G. Research in medical education: balancing service and science. *Adv Health Sci Educ Theory Pract.* 2007;12(1):103-115. DOI: 10.1007/s10459-006-9026-2
18. Regehr G. It's NOT rocket science: rethinking our metaphors for research in health professions education. *Med Educ.* 2010;44(1):31-39. DOI: 10.1111/j.1365-2923.2009.03418.x
19. Whitcomb ME. Research in medical education: what do we know about the link between what doctors are taught and what they do? *Acad Med.* 2002;77(11):1067-1068. DOI: 10.1097/00001888-200211000-00001
20. Gruppen LD. Improving medical education research. *Teach Learn Med.* 2007;19(4):331-335. DOI: 10.1080/10401330701542370
21. Leif M, Semarad N, Ganesan V, Selting K, Burr J, Svec A, Clements P, Talmon G. The Quality Of Evidence In Preclinical Medical Education Literature: A Systematic Review. *Adv Med Educ Pract.* 2019;10:925-933. DOI: 10.2147/AMEP.S212858
22. Colliver JA, McGaghie WC. The reputation of medical education research: quasi-experimentation and unresolved threats to validity. *Teach Learn Med.* 2008;20(2):101-103. DOI: 10.1080/10401330801989497
23. Stokes DE. *Pasteur's Quadrant. Basic Science and Technological Innovation*. Washington, D.C.: Brookings Institution Press; 1997.
24. Jovanović M. *Fußspuren in der Publikationslandschaft: Einordnung wissenschaftlicher Themen und Technologien in grundlagen- und anwendungsorientierte Forschung mithilfe bibliometrischer Methoden*. Stuttgart: Frauenhofer-Verlag; 2011.
25. Hoffmann N. *Dokumentenanalyse in der Bildungs- und Sozialforschung. Überblick und Einführung*. Weinheim: Beltz Juventa; 2017.

26. Flick U. Qualitative Forschung. Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften. 5. Aufl. Reinbek bei Hamburg: Rowohlt; 2000.
27. Wirtz M, Caspar F. Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Göttingen: Hogrefe; 2002.
28. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
29. Brüggemann A, Bromme R. Anwendungsorientierte Grundlagenforschung in der Psychologie: Sicherung von Qualität und Chancen in den Beurteilungs- und Entscheidungsprozessen der DFG. *Psycholog Rundschau*. 2006;57(2):112-118.
30. Döring N, Bortz J. Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. 5. vollst. überarb. aktual. und erwei. Aufl. Berlin: Springer; 2016. DOI: 10.1007/978-3-642-41089-5
31. Reinders H, Gräsel C, Ditton H. Praxisbezug Empirischer Bildungsforschung. In: Reinders H, Ditton H, Gräsel C, Gniewosz B, editors. *Empirische Bildungsforschung. Gegenstandsbereiche*. Wiesbaden: VS Verl. für Sozialwiss. 2011. p.221-233. DOI: 10.1007/978-3-531-93021-3_19
32. Fischer F, Waibel M, Wecker C. Nutzenorientierte Grundlagenforschung im Bildungsbereich. *Zeitschr Erzwiss*. 2005;8(3):427-442.
33. Bush V. Science: The Endless Frontier; A report to the President on a Program for Postwar Scientific Research. Reprint anlässlich des 75. Jahrestages. National Science Foundation; 2020. Zugänglich unter/available from: https://www.nsf.gov/about/history/EndlessFrontier_w.pdf
34. Ellaway RH, Hecker KG. What role does basic research have in an applied field? *Adv Health Sci Educ Theory Pract*. 2022;27(2):289-292. DOI: 10.1007/s10459-022-10117-7
35. Tolsgaard MG. Medical education research in obstetrics and gynecology. *Am J Obst Gyn*. 2019;220(1):121. DOI: 10.1016/j.ajog.2018.09.013
36. Meyer HS, Durning SJ, Sklar DP, Maggio LA. Making the first cut: an analysis of academic medicine editors' reasons for not sending manuscripts out for external peer review. *Acad Med*. 2018;93(3):464-470. DOI: 10.1097/ACM.0000000000001860
37. Prystowsky JB, Bordage G. An outcomes research perspective on medical education: the predominance of trainee assessment and satisfaction. *Med Educ*. 2001;35(4):331-336. DOI: 10.1046/j.1365-2923.2001.00910.x
38. Biesta GJJ, van Braak M. Beyond the medical model: Thinking differently about medical education and medical education research. *Teach Learn Med*. 2020;32(4):449-456. DOI: 10.1080/10401334.2020.1798240
39. Norman G. Editorial - how bad is medical education research anyway? *Adv Health Sci Educ Theory Pract*. 2007;12(1):1-5. DOI: 10.1007/s10459-006-9047-x
40. Lurie SJ. Raising the passing grade for studies of medical education. *JAMA*. 2003;290(9):1210-1212. DOI: 10.1001/jama.290.9.1210
41. Bolander Laksov K, Dornan T, Teunissen PW. Making theory explicit - An analysis of how medical education research(ers) describe how they connect to theory. *BMC Med Educ*. 2017;17(1):18. DOI: 10.1186/s12909-016-0848-1
42. Nichols TR, Wisner, PM, Cripe G, Gulabchand L. Putting the Kappa Statistic to Use. *Qual Assur J*. 2010;13:57-61. DOI: 10.1002/qaj.481
43. Souvignier E, Gold A. Wirksamkeit von Lehrmethoden. In: Schweizer K, editor. *Leistung und Leistungsdiagnostik*. Berlin: Springer; 2006. p.146-166. DOI: 10.1007/3-540-33020-8_10
44. Tenorth HE, Tippelt R, editors. *Lexikon Pädagogik*. Weinheim: Beltz; 2007. p.137f.
45. Döring N. Forschungsmethoden und Evaluation in Sozial- und Humanwissenschaften. 6. Aufl. Berlin: Springer; 2023. DOI: 10.1007/978-3-662-64762-2

Korrespondenzadresse:

PD Dr. Katrin Schüttpelz-Brauns
 Medizinische Fakultät Mannheim der Universität
 Heidelberg, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim,
 Deutschland
katrin.schuettpelz-brauns@medma.uni-heidelberg.de

Bitte zitieren als

Schüttpelz-Brauns K, Schneider A, Fabry G, Matthes J, Himmelbauer M, Buss B, Giesler M. What is medical education research? An analysis and definition of subjects, objectives and types of research based on articles that have undergone a peer review process. *GMS J Med Educ*. 2026;43(1):Doc12.
 DOI: 10.3205/zma001806, URN: urn:nbn:de:0183-zma0018065

Artikel online frei zugänglich unter

<https://doi.org/10.3205/zma001806>

Eingereicht: 25.10.2024

Überarbeitet: 21.02.2025

Angenommen: 16.04.2025

Veröffentlicht: 15.01.2026

Copyright

©2026 Schüttpelz-Brauns et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.