

Conducting Delphi surveys in medical education research

Abstract

Background: Delphi surveys are becoming increasingly important in medical education research, particularly in the development of curricula, assessment instruments, and recommendations for action. However, due to the flexibility and low standardisation of the method, researchers are faced with the challenge of making numerous methodological decisions before and during a Delphi survey. To ensure a structured and targeted approach, careful planning is essential prior to conducting a Delphi study.

Planning Delphi studies: This article describes how to plan Delphi surveys in the following five steps: 1. Suitability and feasibility of the method, 2. Research question and persons involved, 3. Planning of the survey process up to the first round, 4. Evaluation strategies and planning the follow-up rounds, 5. Presentation and dissemination of the results. Each step is structured on the basis of central questions. The most important aspects are summarised in a checklist.

Conclusion: This guide provides researchers with a comprehensive overview of the methodological possibilities and limitations of Delphi surveys, highlighting potential pitfalls. It supports strategic planning and helps researchers to make informed decisions. In the long run, the quality of Delphi studies in medical education research can thus be improved, enabling the method's potential to be realised more effectively.

Keywords: Delphi technique, education research, program development, research design, knowledge discovery

Angelika Homberg¹

1. Medical Faculty Mannheim,
Heidelberg University,
Division for Study and
Teaching Development,
Department of Medical
Education Research,
Mannheim, Germany

1. Introduction

Medical education research is a broad field that includes the development of consensus-based normative guidelines and recommendations [1]. In this context, Delphi surveys are becoming increasingly important, as they facilitate structured communication between experts, ultimately yielding reliable, differentiated results [2], [3]. They are characterised by the following features: Experts are interviewed anonymously using standardised questionnaires, often with open-ended questions to capture arguments, in at least two rounds. The statistical analysis is usually based on descriptive data. From the second round onwards, the experts receive feedback on the previous round's results and can thus reconsider and, if necessary, revise their assessments. The questionnaire can also be adapted based on the respondents' feedback [4], [5], [6].

Delphi surveys can be used for both qualitative and quantitative analyses. They can be adapted to different contexts and thus used in a variety of ways. They are particularly suitable for dealing with complex issues where knowledge is uncertain or in need of interpretation [6]. As the process unfolds, diverse perspectives can be considered and a consensus on a given task can be achieved [7], [8].

Tasks in medical education research include:

- revising or developing curricula (e.g. [9], [10]),
- defining graduate profiles, competencies or learning objectives (e.g. [11], [12], [13]),
- the revision or development of items for assessment tools (e.g. [14], [15], [16]),
- developing outcome-based indicators for assessing learning outcomes (e.g. [17], [18]),
- collecting and evaluating of ideas and influencing factors for further developing innovative areas (e.g. [19], [20]), and
- conducting needs analyses (e.g. [21], [22]).

Delphi surveys are currently conducted using a variety of different concepts, methodological foundations and quality levels [23], [24]. The term *modification* is used inconsistently [7], [25], [26]. Face-to-face meetings are often incorporated into the survey process to increase the likelihood of the results being applied in practice through direct interaction [27], [28]. In this case, the participants are no longer anonymous.

Due to their complexity, Delphi surveys require the entire research process to be carefully planned. The flexibility of the methodology and the lack of standardisation place greater demands on researchers. In his 2014 standard work, "Delphi-Befragungen", Häder describes some important aspects in detail, focusing on social science and

Table 1: Aspects that lead to the application of the Delphi method (based on Linstone & Turoff 1975 [36])

Field	Aspects
Subject of investigation	While this problem cannot be solved using precise analytical techniques, it can benefit from subjective assessments on a collective basis.
People involved	Those who are set to contribute to the investigation have not yet communicated sufficiently with each other.
	The people involved may come from different backgrounds and have different levels of experience and expertise.
	There are more people needed than can effectively interact in a face-to-face exchange.
Communication process	Time and cost make frequent group meetings unfeasible.
	Face-to-face meetings can be made more efficient through the use of supplementary group communication processes.
Group dynamics	Differences of opinion between individual participants are so serious or sensitive that the communication process needs to be moderated.
	Participants must be guaranteed anonymity so that they can respond to the evaluation without external pressure.
	Existing hierarchies prevent everyone from contributing equally to the process.
	The heterogeneity of the participants must be balanced in order to ensure the validity of the results, i.e. it must be avoided that the results are dominated by the number of participants in a particular interest group or by the strength of the personality of individuals ("bandwagon effect").

futurology [29]. The currently existing recommendations for conducting Delphi surveys in the health sciences (e.g. [23], [25], [30], [31]) and in medical education research (e.g. [32], [33], [34]) were published before 2020. Due to technical and epistemological developments, the methodology has become more specialised in recent years. The DFG-funded DeWiss network (Delphi techniques in health and social sciences) has also contributed to this. As well as addressing specific applications, the network considered fundamental questions relating to the Delphi method and conducted an international Delphi study to develop comprehensive reporting guidelines for the health and social sciences [35].

This "how to" guide aims to provide researchers conducting Delphi surveys in medical education with an overview of the possibilities and challenges of this method. This will enable them to identify potential pitfalls early on and produce valid, well-founded results. Researchers are guided through the five-step process using guiding questions. Finally, a checklist summarising all the important points is provided (see attachment 1). Experienced researchers can also use this "how to" for targeted and structured planning.

2. Planning a Delphi survey

2.1. Step 1: Suitability and feasibility of the method

2.1.1. Is the Delphi survey the most appropriate method?

Firstly, the specific problem to be investigated must be clarified. Whether the Delphi method is suitable depends

primarily on the framework conditions surrounding the problem, rather than the question itself [6]. Table 1 lists aspects that support the use of the Delphi method, based on Linstone and Turoff [36].

If none or very few of these aspects apply, consider whether other consensus methods, such as a consensus conference or the nominal group technique, would be more suitable. As another option, a group discussion, expert survey or workshop could be more appropriate [37].

2.1.2. What resources are available?

The complexity of the survey significantly impacts the amount of resources required. However, with good planning, Delphi surveys can be carried out independently by relatively small research groups within a few months. Conducting the survey as part of a doctoral thesis is also feasible [38].

The time resources required are often underestimated. For each survey round, time slots should be scheduled for follow-up recordings of non-respondents and for periods of high workload (e.g. examination periods) and holidays. Around eight to ten weeks per round is realistic. Between survey rounds, the questionnaires must be evaluated, feedback prepared, and the next round's questionnaire developed and pre-tested quickly, in order to maintain participants' motivation and reduce dropout rates.

Financial resources may be required for incentives, workshops (e.g. room rental and travel costs), producing accompanying materials, and licences for survey tools and programmes for evaluating the data [39]. Simple Delphi surveys can use standard survey tools that guarantee participant anonymity and data security (e.g. [40]). For more complex procedures, specialised tools are

worthwhile because they automate processes such as administration, evaluation, and providing feedback on the data. They also provide templates and enable multi-channel surveys. Real-time calculations of group results can be carried out using individual tools as part of a Real-Time Delphi study [41].

2.1.3. When are Delphi variants used?

In recent decades, various methodological variants have been established [8], which override individual typical features of a standard Delphi survey. For example, *Real-Time Delphi* [41] have no distinct Delphi rounds. In *Group Delphi*, the anonymity of the experts is lifted [42] while *Deliberative Delphi* involves citizens who have been trained by experts are involved instead of the experts themselves [43]. Using a Real-Time or Group Delphi can significantly reduce the required time. In medical education research, the Delphi method is particularly useful for achieving consensus on complex issues through synchronous discussion. Deliberative Delphi studies involving patients (e.g., [44]) or students (e.g., [11], [45], [46]) can also be useful for addressing specific concerns. The publication by Niederberger and Deckert provides a good overview of the characteristics of the currently established variants [8].

2.2. Step 2: Research question and persons involved

2.2.1. Who is involved in planning and managing the process?

To make well-founded and objective decisions associated with a Delphi survey, a steering group should be formed. Typically, steering group members do not participate in the Delphi survey. One exception is the participatory approach, in which steering group members can also form the expert panel (e.g. [44]).

In general, the following people can be considered for the steering group:

- Researchers conducting the Delphi survey.
- Researchers who have experience with the Delphi method.
- Researchers who are familiar with the background of the problem to be investigated.
- Decision-makers, stakeholders, and others, such as deans of studies (strategic approach).
- Those affected, such as students or patients (participatory approach).

Using a strategic or participatory approach can increase the acceptance of the results or their adaptation to specific framework conditions.

It can be useful to involve auxiliary staff to help manage the multiple tasks [29]. It is important to clearly communicate the areas of responsibility and task allocation within the steering group and to consistently document

the planning and the decisions made throughout the process.

2.2.2. How is the research question determined?

The first task of the steering group is to review the research objective and formulate a precise research question. This ensures that all members have the same understanding of the research subject and objective, preventing misunderstandings.

The following should be considered:

- whether technical terms are used and understood equally by all participants;
- whether the results should only be applicable to a specific geographic region, nationally or internationally;
- under what conditions the results are to be applied (e.g. in schools, universities or other educational institutions).
- The overall methodological objective to be achieved should also be considered (e.g. consensus building, idea aggregation or collection of expert opinions).

In medical education research, Delphi surveys are the predominant method for building consensus. The steering group should verify whether other objectives, such as idea aggregation or the collection of expert opinions, are prioritised. Häder provides a good overview of the respective differentiation [29].

2.2.3. Which groups of people are surveyed?

In Delphi studies, respondents are usually referred to as experts: informed individuals or specialists with knowledge in a specific field [25]. As reflective practitioners, they contribute their real-life experiences to the Delphi process [27]. The composition of the expert panel largely depends on the methodological objective of the Delphi survey [29].

In consensus Delphi studies, the population diversity of participants who have relevant expertise on the research topic should be represented. All relevant groups should be integrated into the Delphi process in the most balanced way possible [47], [48] (see table 2), with each group represented by at least two people, preferably three. In this way, different points of view within the respective group can be recorded and a dropout of individual persons has less influence on the overall result. In principle, additional experts can be recruited later on. This should be considered, in particular, if dropping out could lead to distorted results (e.g., if a disproportionate number of experts leave a group) or reduce statistical significance due to a smaller sample size (e.g., in small samples). In order to maintain the comparability, consistency and integrity of the study, the new participants should have comparable expertise and be familiarised with the previous results and the course of the study. The researchers must disclose and justify the subsequent recruitment.

Table 2: Balanced composition of the expert panel depending on the objective of the Delphi survey

Selection criteria	Example
Professional group	Interprofessional studies: Doctors, nurses and therapists
Region or nation	International studies: Participants from infrastructurally strong and weak countries
Hierarchical level	Development of criteria for clinical teaching: Senior physicians and assistant physicians
Field of experience	Development of graduate profiles: Experts from the outpatient and inpatient sector
Perspective	Development of longitudinal curricula for clinical skills: Students from different phases of training, lecturers, patients
Socio-demographic data, such as gender, age and institutional affiliation	Development of gender-appropriate learning materials: Female and male students

One of the major difficulties is defining what constitutes an expert [49]. The steering group should therefore discuss this issue in relation to the objectives of the Delphi survey, since the composition of the expert panel has a significant influence on the reliability and validity of the study.

2.2.4. How is the sample size determined?

As complete enumeration is generally not possible with Delphi surveys, a sample must be selected. Two factors are particularly relevant to the success of the survey: the respondents' individual expertise in relation to the task and their personal interest in the topic. Targeted selection can significantly reduce dropout rates [50]. The following criteria can be used to proceed systematically: Number of published articles on the subject area in question or publication index, participation in specialist committees, professional role or activities, active participation in relevant networks and forums. Faculties or universities may also be asked to nominate experts in specific subject areas. Students can be recruited via the student council, for example, and patients via self-help groups. The personal contacts and affiliations of the steering group members with relevant networks can be helpful in identifying motivated individuals.

2.2.5. How many experts should be included?

There are no uniform recommendations on how many experts should be included. According to Turoff [51], a group of ten to forty people is favorable. Trevelyan et al. suggest eight to fifteen participants for homogeneous panels in the field of health sciences [30]. Delphi surveys in medical education research use panels ranging in size from eight participants (e.g. [38]) to over one hundred participants (e.g. [52]).

The time required increases with the number of participants, especially if the survey contains a high percentage of qualitative questions and the steps are not automated.

A small expert panel is sufficient under the following conditions:

- Only a few topics, theses and questions need to be dealt with.
- Only a few different expert opinions, viewpoints and perspectives need to be taken into account.
- There is a small population that can be mapped.
- The Delphi survey can be carried out in a short period of time.
- The motivation of the participants, and thus the expected response rate, is estimated to be rather high.
- The expected willingness of participants to contribute arguments and ideas is considered to be high.
- It is possible to recruit replacement experts after the first round.

2.3. Step 3: Planning the survey process up to the first round

2.3.1. How are the items for the first questionnaire developed?

In the field of educational research, Garavalia et al. [53], following Linstone [36], describe the classic Delphi procedure as consisting of three rounds.

- *First round:* Collection of items based on an open-ended question.
- *Second round:* Evaluation of the collected items using rating scales and arguments captured through open-ended questions.
- *Third round:* Re-evaluation of the items based on feedback comments and arguments [53].

This procedure is rarely used in medical education research (e.g. [11], [12]). More often, items are formulated in advance. The participating experts are asked to evaluate these items and suggest further ones. This can speed up the Delphi process. Additionally, the experts gain an understanding of what is expected of them and are more likely to generate comparable new content. This reduces both the workload for the respondents and the likelihood of them deviating from the topic. The development of the items included in the first survey round depends on the available resources [29]. For certain questions, statements can be taken from lists compiled by professional

associations or committees. Often, relevant publications are available, enabling the steering group to compile the items based on the literature. If these requirements are not met, the steering group can develop the items directly. In this case, it is advisable to consult additional experts from specialist committees and bodies, either in person before the Delphi survey or via a preliminary survey.

2.3.2. How is the number of rounds determined?

This should be decided at the start of the study, regardless of whether a consensus has been reached on all items. If achieving consensus is defined as a stopping criterion, the survey will continue until all items have been clearly assigned. Therefore, the effort involved cannot be calculated for either the steering group or the participants. The most important factors to consider when deciding how many rounds there will be are: how well the topic has already been researched, how heterogeneous the panel of experts is, how the consensus is defined, what result is to be achieved, and what resources are available. For example, Ellaway et al. [10] conducted a Delphi study to develop a curriculum for medical students on providing care to transgender and gender-diverse patients. The complexity of the topic required four rounds. However, good results can often be achieved in just two rounds (e.g. [54], [55], [56]).

2.3.3. How is consensus measured?

In Delphi surveys, consensus means that a certain level of agreement is reached among the experts surveyed on a statement. However, it does not imply that all experts share the same opinion. In most publications, consensus is defined as being reached when 70-85% of experts agree with a statement [57]. Agreement is usually measured using descriptive statistical measures and depends on the scale level [58]. At a nominal scale level, the relative frequency of mentions can be determined. Typically, the degree of agreement is measured on an ordinal scale. Here, too, it is possible to determine the proportion of respondents who agree or disagree with or reject a statement once a certain characteristic threshold is reached. The median and interquartile range (IQR) are used more frequently because they allow majorities to be depicted [59]. For example, in a 5-point rating scale, an IQR <1 can be defined as consensus [58]. The mean and standard deviation should be avoided for ordinal scales. Consensus can also be determined using ranking questions. However, it should be noted that, despite requiring more cognitive effort from respondents, ranking questions generally offer no additional benefit over rating questions [60].

2.3.4. How should the first questionnaire be structured?

The first Delphi round questionnaire should open with a preamble and collect demographic data. It should also contain a central section comprising both open-ended and closed-ended questions. The scale levels and options for the closed-ended questions must be chosen carefully [28], [59], [61] (see table 3).

A clear structure to the first questionnaire and an easily manageable amount of collected data are key to ensuring that follow-up rounds and feedback are not overloaded and can be appropriately designed [30]. Belton et al. recommend that the processing time for a Delphi round should not exceed 30 minutes, and that the first round should only include a few open-ended questions [62].

2.3.5. How should communication with experts be handled?

It is important to treat the participating experts as partners and to explain all important steps. The process begins with clearly identifying the area of expertise. This means that interviewees should not be referred to as experts until the term has been defined [30]. The request for participation should be personalised. Explaining why their expertise is important increases their willingness to participate. To maintain motivation, social rewards are often more effective than financial ones [62]. For example, participants may be motivated if they recognise that their answers influence the outcome of the Delphi survey, or if they are provided with background information and interim results. Furthermore, using the appropriate language in the questionnaire and accompanying materials is crucial for success. In the case of heterogeneous expert panels, such as those in international studies or participatory Delphi surveys, it is particularly important to involve representatives from each subgroup of the expert panel in the pretest, to ensure the questionnaire is comprehensible to all participants [63].

2.4. Step 4: Evaluation strategies and planning the follow-up rounds

2.4.1. Which strategies are used for the analysis?

The evaluation strategy depends on the type of data available. Quantitative data is typically analysed using descriptive statistics. For consensus measurement, each item is considered individually and placed into groups based on defined consensus criteria: *Consensus of agreement, dissent, or consensus of disagreement*.

Responses to open-ended questions are analysed according to their function [64] (see table 3). Exploratory questions can be categorised and summarised using either quantitative content analysis [65] or thematic analysis [66]. Questions on extension can be analysed using

Table 3: Structure of the questionnaire for the first Delphi round

Chapter	Content	Comment
Preamble	- Aim and context of the survey - Use of key terms - Structure of the questionnaire - Processing time - Process of preparing and providing feedback on the results - Overall survey process	If the respondents require extensive background information or many technical terms need to be explained, these can be sent by e-mail as accompanying material or glossary or linked directly in the questionnaire (mouse over).
Recording demographic data and conflicts of interest	Characteristics enabling the transparent presentation of the expertise, perspectives and possible conflicts of interest of the expert panel	If only fully completed questionnaires are evaluated, this block can also be placed at the end of the survey.
Main part: closed-ended questions	- Presentation of the individual statements or items as theses - Evaluation on a 3 to 9-point rating scale plus the option "no answer"	The number of levels increases the evaluation options for researchers and the cognitive effort required of respondents. Having a middle option could counteract the distortion of the results.
Main part: open-ended questions	Open-ended questions in combination with theses: - Collection of arguments for an assessment (<i>argumentative questions</i>) - Collection of further items (<i>questions on extension</i>) Stand-alone questions - Collecting ideas on a specific issue (<i>exploratory questions</i>)	Advantage of open-ended questions: Respondents can answer in their own words, enabling researchers to gain valuable insights into their motivations. Disadvantage of open-ended questions: They require more cognitive effort and time from respondents, who may be less willing to answer.

quantitative content analysis [67]. Evaluating qualitatively recorded arguments in which interviewees justify their quantitative evaluations plays a special role. The five-stage, *Argument-based QUALitative Analysis* (AQUA) method was developed specifically for this purpose. This method enables qualitative data to be evaluated in a structured way and prepared for feedback alongside quantitative data [68].

2.4.2. How is feedback transmitted?

After each Delphi round, the participants should receive an objective feedback report on the aggregated group results. The group composition should be presented anonymously. Quantitative data can be visualised using position and scatter measures, e.g., bar or column charts. Individualised feedback, where each participant's evaluation is compared to the group values, is usually time-consuming. Its usefulness is often questioned, as it creates pressure to conform and increases cognitive load [69].

The qualitative results should be well structured and remain close to the original wording, enabling the experts to reflect on their contributions. The aim is to encourage participants to reconsider their views in light of the group

responses and potentially adjust them. Figure 1 illustrates how quantitative data can be presented concisely alongside individual feedback and arguments.

The feedback report can either be sent to respondents as a separate document alongside the invitation to the follow-up round or integrated into the follow-up questionnaire.

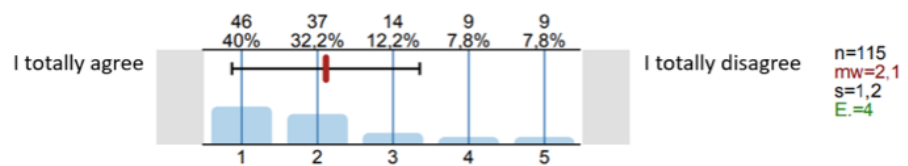
2.4.3. How is the determined consensus dealt with?

Items for which a consensus of agreement or disagreement has been reached are presented in the interim and final reports, but will not be re-evaluated in the follow-up round. Although a new vote could determine the reproducibility and stability of the ratings, the additional burden on respondents usually outweighs this advantage.

Items for which no consensus was reached regarding agreement or disagreement are analysed [70]. Depending on the outcome of this analysis, the items may be terminated, modified or resubmitted unchanged alongside the evaluation arguments [36] (see figure 2).

Item 4.3: Open-book exams should be used more frequently for assessment purposes

Statistics



Arguments

Agreement	Disagreement
- Competence-oriented testing is encouraged - Question design is more flexible	- The correction effort is too high - Assessments are difficult to quantify

Figure 1: Example of feedback design in Delphi surveys

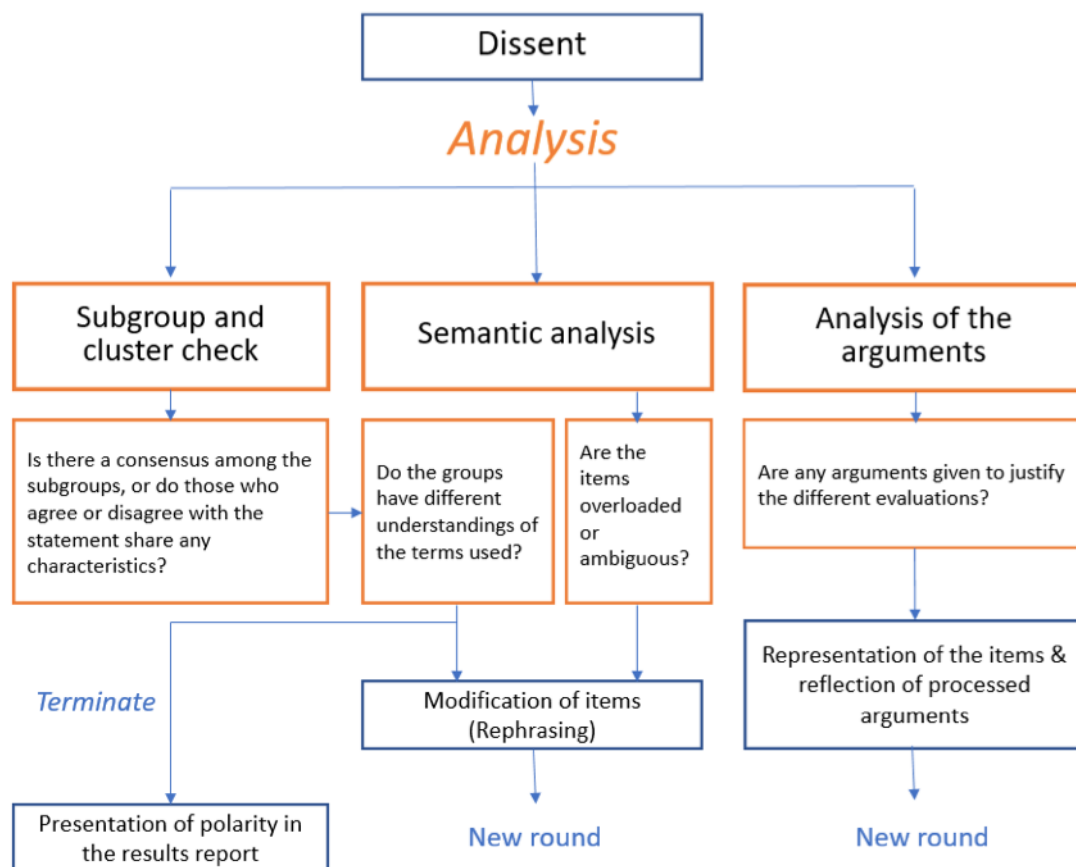


Figure 2: Analysis techniques for items that remain in disagreement in Delphi surveys

2.4.4. How are the follow-up questionnaires structured?

Like the initial questionnaire, the follow-up questionnaires contain identical or modified items from the previous round. They also include items generated from exploratory questions. Furthermore, experts should have the opportunity to justify their judgements. Open-ended exploratory questions should not be included in the further course of the Delphi survey, or only to a limited extent.

2.5. Step 5: Presentation and dissemination of the results

2.5.1. What about items that remain in dissent?

Even after many rounds, it is possible that no consensus has been reached on individual items. This must be addressed and analysed.

For example, it could be that the experts:

- generally have different priorities for values or goals,
- weight or consider partial aspects differently, or
- have conflicts of interest and therefore cannot deviate from their positions.

The cause may also lie in the methodological implementation. For example:

- the topic area may not have been narrowed down sufficiently in advance,
- arguments were not communicated or taken into account sufficiently in the feedback,
- the number of rounds was insufficient, or
- the items may not have been formulated precisely during the process.

Any remaining disagreements can be discussed further within the steering group, in a final meeting with the participating experts, or with the involvement of other experts or stakeholders. In any case, dissenting opinions should be included in the results report.

2.5.2. How are the results published?

In order to increase the quality and transparency of reporting, a corresponding reporting guideline should be used. DelphiSTAR (Delphi studies in social and health sciences - Recommendations for an interdisciplinary standardized reporting) was developed specifically for publications in the health and social sciences [35]. Alternatively, one could use ACCORD (ACcurate CONsensus Reporting Document), developed for biomedicine, or CREDES (Guidance on Conducting and Reporting DELphi Studies), developed for the palliative care sector [31]. Participating experts contribute their experience and a considerable amount of time to the Delphi process. This should be acknowledged, and, if consent has been given, the experts should be named. Their qualifications can be

presented in an attachment to illustrate their contributions to the Delphi process or to disclose conflicts of interest.

2.5.3. How are the results disseminated and implemented?

Even before beginning the study, researchers should consider whether measures are necessary to disseminate and implement the results. Often, the results must be processed further before they can be put into practice. For instance, summaries can be created and distributed to target groups. A final workshop can also be helpful for this purpose (e.g. [71], [72]). Depending on the outcome, it may be useful to evaluate the results' feasibility in practice and examine further application and implementation possibilities.

3. Discussion

This article provides a concise guide to conducting Delphi surveys in medical education research. It outlines key steps, explains standard procedures and alternatives, and highlights potential pitfalls. Readers are expected to have a basic understanding of the principles of empirical social research, including the construction of questionnaires, the operationalisation of theoretical concepts, and the analysis and interpretation of data. Researchers wishing to use Delphi variants must familiarise themselves with the relevant in-depth literature.

It is important to remember that methodological decisions often have far-reaching consequences. For instance, the composition of the expert panel [29], the choice of rating scale, and how consensus is determined [59] can all significantly impact the results. The statistical presentation of the results may mislead people into interpreting them as objective measurements and overestimating their significance. In medical education research, however, they are better understood as the result of a negotiation process in a given situation, whereby an agreement is reached on how something should or could be done. A major challenge is that many issues relating to Delphi surveys remain unresolved. For example, there is a lack of secondary analyses, experimental studies, or methodological tests that would allow for well-founded conclusions to be drawn about the influence of methodological and statistical procedures on judgement behaviour and, consequently, on consensus building [3], [58]. Nevertheless, this method is the preferred way to systematically bring together the specialised knowledge and real-life experience of experts from science and practice to increase certainty in planning and action. This guide can help to improve the quality of Delphi studies in medical education research. The steps and recommendations mentioned here can largely be transferred to other Delphi studies in health and educational sciences.

Author's ORCID

Angelika Homberg: [0000-0001-5585-1126]

Competing interests

The author declares that she has no competing interests.

Attachments

Available from <https://doi.org/10.3205/zma001800>

- Attachment_1.pdf (130 KB)
Checklist for conducting Delphi studies in medical education research

References

- 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/zma001804
- Foth T, Efstathiou N, Vanderspank-Wright B, Ufholz LA, Dutthorn N, Zimansky M, Humphrey-Murto S. The use of Delphi and Nominal Group Technique in nursing education: A review. *Int J Nurs Stud.* 2016;60:112-120. DOI: 10.1016/j.ijnurstu.2016.04.015
- Humphrey-Murto S, Varpio L, Wood TJ, Gonsalves C, Ufholz LA, Mascioli K, Wang C, Foth T. The Use of the Delphi and Other Consensus Group Methods in Medical Education Research: A Review. *Acad Med.* 2017;92(10):1491-1498. DOI: 10.1097/acm.0000000000001812
- Webler T, Levine D, Rakel H, Renn O. A novel approach to reducing uncertainty: The group Delphi. *Technol Forecast Soc Change.* 1991;39(3):253-263. doi: 10.1016/0040-1625(91)90040-M
- Niederberger M, Renn O, editors. *Delphi-Verfahren in den Sozial- und Gesundheitswissenschaften: Konzept, Varianten und Anwendungsbeispiele.* Wiesbaden: Springer VS; 2019. DOI: 10.1007/978-3-658-21657-3
- Linstone HA, Turoff M. *The Delphi Method: Techniques and Applications.* Boston: Addison-Wesley Publishing Company; 2002.
- Homberg A, Schüttpelz-Brauns K. Delphi-Befragungen in der Ausbildungsforschung – Anwendungsbereiche und Umsetzung. In: *Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA) und des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ).* Halle (Saale), 15.-17.09.2022. Düsseldorf: German Medical Science GMS Publishing House; 2022. DocV-10-04. DOI: 10.3205/22gma062
- Niederberger M, Deckert S. Das Delphi-Verfahren: Methodik, Varianten und Anwendungsbeispiele [The Delphi technique: Methodology, variants and usage examples]. *Z Evid Fortbild Qual Gesundhwes.* 2022;174:11-19. DOI: 10.1016/j.zefq.2022.08.007
- Gordon M, Baker P, Catchpole K, Darbyshire D, Schocken D. Devising a consensus definition and framework for non-technical skills in healthcare to support educational design: A modified Delphi study. *Med Teach.* 2015;37(6):572-577. DOI: 10.3109/0142159X.2014.959910
- Ellaway RH, Thompson NL, Temple-Oberle C, Pacaud D, Frecker H, Jablonski TJ, Demers J, Mattatall F, Raiche J, Hull A, Jailil R. An undergraduate medical curriculum framework for providing care to transgender and gender diverse patients: A modified Delphi study. *Perspect Med Educ.* 2022;11(1):36-44. DOI: 10.1007/s40037-021-00692-7
- Berger-Estilita J, Nabecker S, Greif R. A Delphi consensus study for teaching "Basic Trauma Management" to third-year medical students. *Scand J Trauma Resusc Emerg Med.* 2019;27(1):91. DOI: 10.1186/s13049-019-0675-6
- Almeland SK, Lindford A, Berg JO, Hansson E. A core undergraduate curriculum in plastic surgery - a Delphi consensus study in Scandinavia. *J Plast Surg Hand Surg.* 2018;52(2):97-105. DOI: 10.1080/2000656X.2017.1343190
- Behrend R, Herinek D, Kienle R, Arnold F, Peters H. Entwicklung interprofessioneller Ausbildungsziele für die Gesundheitsberufe an der Charité – Universitätsmedizin Berlin – Eine Delphi-Studie [Development of Interprofessional Learning Outcomes for Health Professions at Charité - Universitätsmedizin Berlin - A Delphi-Study]. *Gesundheitswesen.* 2022;84(6):532-538. DOI: 10.1055/a-1341-1368
- Al Asmri M, Haque MS, Parle J. A Modified Medical Education Research Study Quality Instrument (MMERSQI) developed by Delphi consensus. *BMC Med Educ.* 2023;23(1):63. DOI: 10.1186/s12909-023-04033-6
- Fluit C, Bolhuis S, Grol R, Ham M, Feskens R, Laan R, Wensing M. Evaluation and feedback for effective clinical teaching in postgraduate medical education: Validation of an assessment instrument incorporating the CanMEDS roles. *Med Teach.* 2012;34(11):893-901. DOI: 10.3109/0142159X.2012.699114
- Kahr Rasmussen N, Nayahangan LJ, Carlsen J, Ekberg O, Brabrand K, Albrecht-Beste E, Bachmann Nielsen M, Konge L. Evaluation of competence in ultrasound-guided procedures-a generic assessment tool developed through the Delphi method. *Eur Radiol.* 2021;31(6):4203-4211. DOI: 10.1007/s00330-020-07280-z
- Rau T, Fegert JM, Kornmann M, Muche R, Gulich M, Barth H, Kornmann M, Fegert JM, Öchsner W. Erarbeitung von Empfehlungen zu universitären Prüfungen an der Medizinischen Fakultät Ulm. *GMS Z Med Ausbild.* 2010;27(1):Doc07. DOI: 10.3205/zma000644
- de Leeuw RA, Walsh K, Westerman M, Scheele F. Consensus on Quality Indicators of Postgraduate Medical E-Learning: Delphi Study. *JMIR Med Educ.* 2018;4(1):e13. DOI: 10.2196/mededu.9365
- Homberg A, Stock-Schröer B. Interprofessional Education on Complementary and Integrative Medicine. *Clin Teach.* 2021;18(2):152-157. DOI: 10.1111/tct.13280
- Teo YH, Peh TY, Abdurrahman A, Lee AS, Chiam M, Fong W, Wijaya L, Krishna LK. A modified Delphi approach to enhance nurturing of professionalism in postgraduate medical education in Singapore. *Singapore Med J.* 2024;65(6):313-325. DOI: 10.11622/smedj.2021224
- Thim S, Nayahangan LJ, Paltved C, Jensen RD, Konge L, Hertel NT, Balslev T. Identifying and prioritising technical procedures for simulation-based curriculum in paediatrics: a Delphi-based general needs assessment. *BMJ Paediatr Open.* 2020;4(1):e000697. DOI: 10.1136/bmjpo-2020-000697
- Bessmann EL, Østergaard HT, Nielsen BU, Russell L, Paltved C, Østergaard D, Konge L, Nayahangan LJ. Consensus on technical procedures for simulation-based training in anaesthesiology: A Delphi-based general needs assessment. *Acta Anaesthesiol Scand.* 2019;63(6):720-729. DOI: 10.1111/aas.13344

23. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS One*. 2011;6(6):e20476. DOI: 10.1371/journal.pone.0020476
24. Schifano J, Homberg A, Sonnberger M, Niederberger M. Reporting guidelines for Delphi techniques in health sciences: A methodological review. *Z Evid Fortbild Qual Gesundhwes*. 2022;172:1-11. DOI: 10.1016/j.zefq.2022.04.025
25. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. *Int J Nurs Stud*. 2001;38(2):195-200. DOI: 10.1016/s0020-7489(00)00044-4
26. Schifano J, Niederberger M. How Delphi studies in the health sciences find consensus: a scoping review. *Syst Rev*. 2025;14(1):14. DOI: 10.1186/s13643-024-02738-3
27. Homberg A, Hirt J, Niederberger N, Häder M, Vollmar H, Cuhls K, Reiber K, Mohr J. Kann Delphi nur Zukunftsforschung? – Ein Blick in die Gesundheitswissenschaften. *Z Zukunftsforsch*. 2024;12(1). Zugänglich unter/available from: <https://www.zeitschrift-zukunftsforschung.de/zukunftsforschung/article/view/7>
28. Shang Z. Use of Delphi in health sciences research: A narrative review. *Medicine (Baltimore)*. 2023;102(7):e32829. DOI: 10.1097/MD.00000000000032829
29. Häder M. Delphi-Befragungen: Ein Arbeitsbuch. 3. Auflage. Wiesbaden: Springer VS; 2014. DOI: 10.1007/978-3-658-01928-0
30. Trevelyan EG, Robinson PN. Delphi methodology in health research: how to do it? *Eur J Integr Med*. 2015;7(4):423-428. DOI: 10.1016/j.eujim.2015.07.002
31. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on Conducting and Reporting DELphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliat Med*. 2017;31(8):684-706. DOI: 10.1177/0269216317690685
32. Waggoner J, Carline JD, Durning SJ. Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research. *Acad Med*. 2016;91(5):663-668. DOI: 10.1097/ACM.0000000000001092
33. Donohoe H, Stellefson M, Tennant B. Advantages and Limitations of the e-Delphi Technique. *Am J Health Educ*. 2012;43(1):38-46. DOI: 10.1080/19325037.2012.10599216
34. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and Nominal Group in medical education research. *Med Teach*. 2017;39(1):14-19. DOI: 10.1080/0142159x.2017.1245856
35. Niederberger M, Schifano J, Deckert S, Hirt J, Homberg A, Köberich S, Kuhn R, Rommel A, Sonnberger M; DEWISS network. Delphi studies in social and health sciences-Recommendations for an interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. *PLoS One*. 2024;19(8):e0304651. DOI: 10.1371/journal.pone.0304651
36. Linstone HA, Turoff M. The Delphi method: Techniques and applications. Addison-Wesley: Reading, Mass; 1975.
37. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm*. 2016;38(3):655-662. DOI: 10.1007/s11096-016-0257-x
38. Shewade HD, Jeyashree K, Kalaiselvi S, Palanivel C, Panigrahi KC. Competency-based tool for evaluation of community-based training in undergraduate medical education in India - a Delphi approach. *Adv Med Educ Pract*. 2017;8:277-286. DOI: 10.2147/AMEP.S123840
39. Niederberger M, Köberich S; members of the DeWiss Network. Coming to consensus: the Delphi technique. *Eur J Cardiovasc Nurs*. 2021;20(7):692-695. DOI: 10.1093/eurjcn/zvab059
40. Spranger J, Niederberger M. Handbuch: Unipark für Delphi-Befragungen. Schwäbisch Gmünd: University of Education; 2023. Zugänglich unter/available from: https://phsg.bs-zw.de/frontdoor/deliver/index/docId/455/file/Handbuch_unipark_Delphi-Befragungen_SoSe23.pdf
41. Aengenheyster S, Cuhls K, Gerhold L, Heiskanen-Schüttler M, Huck J, Muszynska M. Real-Time Delphi in practice – A comparative analysis of existing software-based tools. *Technol Forecast Soc Change*. 2017;118:15-27. DOI: 10.1016/j.techfore.2017.01.023
42. Niederberger M, Renn O. Das Gruppendelphi-Verfahren. Wiesbaden: Springer Fachmedien; 2018. DOI: 10.1007/978-3-658-18755-2
43. Groetker R. Expertenkonsultationen und Stakeholder-Befragungen mit Deliberativem Delphi (Expert Consultations and Stakeholder-Surveys with Deliberative Delphi). *SSRN Electronic J*. 27.09.2017. DOI: 10.2139/ssrn.3256258
44. Estrella Porter PD, Ayala Mullo JF, Barba Carrera DA, Barros Castro AX, Cabascango Vasquez ES, Del Castillo Arellano JC, Condo Espinel PX, Eid Arellano EG, Estrella Porter JA, Falconi Paez AC, Fierro Valle MC, Gallegos Miranda PJ, Guerra Velastegui AR, Guevaraz Baez DC, Jara Santamaria BD, Lopez Diaz JA, Mejia Viana JC, Moya Quitto GF, Muenala PM, Muenala TC, Nocolalade López Bi, Oquendo Carrera AD, Ordoñez Paz AL, Ortiz Duque EF, Palacios Granda MC, Patoja Borja NS, Puga Martinez SE, Rueda Ordoñez CJ, Soto Gutierrez LP, Tixi Tapia GE, Torres Moscoso MB, Vaca Porras MA, Viteri Suárez MI, Gullemot JR. Medical education from the point of view of medical students: Results from four participatory Delphi panels in Quito, Ecuador. *Med Teach*. 2020;42(9):1051-1057. DOI: 10.1080/0142159X.2020.1792865
45. Copeland C, Fisher J, Teodorczuk A. Development of an international undergraduate curriculum for delirium using a modified delphi process. *Age Ageing*. 2018;47(1):131-137. DOI: 10.1093/ageing/afx133
46. Salihu HM, Dongarwar D, Malmberg ED, Harris TB, Christner JG, Thomson WA. Curriculum Enrichment Across the Medical Education Continuum Using e-Delphi and the Community Priority Index. *South Med J*. 2019;112(11):571-580. DOI: 10.14423/SMJ.0000000000001033
47. Duncan E, O'Cathain A, Rousseau N, Croot L, Sworn K, Turner KM, Yardley L, Hoddinott P. Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. *BMJ Open*. 2020;10(4):e033516. doi: 10.1136/bmjopen-2019-033516
48. Spickermann A, Zimmermann M, von der Gracht HA. Surface- and deep-level diversity in panel selection – Exploring diversity effects on response behaviour in foresight. *Technol Forecast Soc Change*. 2014;85:105-120. DOI: 10.1016/j.techfore.2013.04.009
49. Baker J, Lovell K, Harris N. How expert are the experts? An exploration of the concept of 'expert' within Delphi panel techniques. *Nurse Res*. 2006;14(1):59-70. DOI: 10.7748/nr2006.10.14.1.59.c6010
50. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs*. 2000;32(4):1008-1015.
51. Turoff M. The Policy Delphi. In: Linstone HA, Turoff M, editors. *The Delphi method: Techniques and applications*. Newark (NJ): New Jersey Institute of Technology; 2002. p.80-96.
52. Viljoen CA, Millar RS, Manning K, Burch VC. Determining electrocardiography training priorities for medical students using a modified Delphi method. *BMC Med Educ*. 2020;20(1):431. DOI: 10.1186/s12909-020-02354-4

53. Garavalia L, Gredler M. Teaching Evaluation through Modeling: Using the Delphi Technique to Assess Problems in Academic Programs. *Am J Eval*. 2004;25(4):375-380. DOI: 10.1016/j.ameval.2004.05.006
54. Lei L, Lin L, Deng J, Dong H, Luo Y. Developing an Organ Donation Curriculum for Medical Undergraduates in China Based on Theory of Planned Behavior: A Delphi Method Study. *Ann Transplant*. 2020;25:e922809. DOI: 10.12659/AOT.922809
55. Shah S, McCann M, Yu C. Developing a National Competency-Based Diabetes Curriculum in Undergraduate Medical Education: A Delphi Study. *Can J Diabetes*. 2020;44(1):30-36.e2. DOI: 10.1016/j.cjcd.2019.04.019
56. Urushibara-Miyachi Y, Kikusawa M, Ikusaka M, Otaki J, Nishigori H. Lists of potential diagnoses that final-year medical students need to consider: a modified Delphi study. *BMC Med Educ*. 2021;21(1):234. DOI: 10.1186/s12909-021-02652-5
57. Barrios M, Guileria G, Nuño L, Gómez-Benito J. Consensus in the delphi method: What makes a decision change? *Technol Forecast Soc Change*. 2021;163:120484. DOI: 10.1016/j.techfore.2020.120484
58. von der Gracht HA. Consensus measurement in Delphi studies. *Technol Forecast Soc Change*. 2012;79(8):1525-1536. DOI: 10.1016/j.techfore.2012.04.013
59. Lange T, Kopkow C, Lutzner J, Gunther KP, Gravius S, Scharf HP, Stöve J, Wagner R, Schmitt J. Comparison of different rating scales for the use in Delphi studies: different scales lead to different consensus and show different test-retest reliability. *BMC Med Res Methodol*. 2020;20(1):28. DOI: 10.1186/s12874-020-0912-8
60. Del Grande C, Kaczorowski J. Rating versus ranking in a Delphi survey: a randomized controlled trial. *Trials*. 2023;24(1):543. DOI: 10.1186/s13063-023-07442-6
61. Johns R. One Size Doesn't Fit All: Selecting Response Scales For Attitude Items. *J Elect Public Opin Parties*. 2005;15(2):237-264. DOI: 10.1080/13689880500178849
62. Belton I, MacDonald A, Wright G, Hamlin I. Improving the practical application of the Delphi method in group-based judgement: a six-step prescription for a well-founded and defensible process. *Health Technol Assess*. 2019;147:72-82. DOI: 10.1016/j.techfore.2019.07.002
63. Artino Jr AR, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach*. 2014;36(6):463-474. DOI: 10.3109/0142159x.2014.889814
64. O'Cathain A, Thomas KJ. "Any other comments?" Open questions on questionnaires - a bane or a bonus to research? *BMC Med Res Methodol*. 2004;4:25. DOI: 10.1186/1471-2288-4-25
65. Mayring P. Einführung in die qualitative Sozialforschung : eine Anleitung zu qualitativem Denken. 7., überarbeitete Auflage. Weinheim, Basel: Beltz; 2023.
66. Braun V, Clarke V. Thematic analysis: a practical guide. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE; 2022.
67. Fröh W. Inhaltsanalyse: Theorie und Praxis. 9. überarbeitete Auflage. Konstanz: UVK Verlagsgesellschaft mbH; 2017. DOI: 10.36198/9783838547350
68. Niederberger M, Schifano J, Deckert S, Hirt J, Homberg A, Köberich S, Kuhn R, Rommel A, Sonnberger M; DEWISS network. Delphi studies in social and health sciences-Recommendations for an interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. *PLoS One*. 2024;19(8):e0304651. DOI: 10.1371/journal.pone.0304651
69. Schmidt RC. Managing Delphi Surveys Using Nonparametric Statistical Techniques. *Decis Sci*. 2007;28(3):763-774. DOI: 10.1111/j.1540-5915.1997.tb01330.x
70. Dajani JS, Sincoff MZ, Talley WK. Stability and agreement criteria for the termination of Delphi Studies. *Technol Forecast Soc Change*. 1979;13:83-90. DOI: 10.1016/0040-1625(79)90007-6
71. Gattrell WT, Logullo P, van Zuuren EJ, Price A, Hughes EL, Blazey P, Winchester CC, Tovey D, Goldman K, Hungin AP, Harison N. ACCORD (ACcurate COnsensus Reporting Document): A reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLoS Med*. 2024;21(1):e1004326. DOI: 10.1371/journal.pmed.1004326
72. Wallace MJ, Zecharia A, Guiding C, Tucker S, McFadzean I. Developing a new undergraduate pharmacology core curriculum: The British Pharmacological Society Delphi Method. *Pharmacol Res Perspect*. 2021;9(4):e00832. DOI: 10.1002/prp2.832

Corresponding author:

Angelika Homberg
Medical Faculty Mannheim, Heidelberg University, Division for Study and Teaching Development, Department of Medical Education Research, Theodor-Kutzer-Ufer 1-3, D-68167 Mannheim, Germany
Angelika.homberg@medma.uni-heidelberg.de

Please cite as

Homberg A. Conducting Delphi surveys in medical education research. *GMS J Med Educ*. 2026;43(1):Doc6.
DOI: 10.3205/zma001800, URN: urn:nbn:de:0183-zma0018008

This article is freely available from

<https://doi.org/10.3205/zma001800>

Received: 2024-09-30

Revised: 2025-04-29

Accepted: 2025-06-05

Published: 2026-01-15

Copyright

©2026 Homberg. 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/>.

Durchführung von Delphi-Befragungen in der medizinischen Ausbildungsforschung

Zusammenfassung

Hintergrund: Delphi-Befragungen gewinnen in der medizinischen Ausbildungsforschung an Bedeutung, insbesondere für die Entwicklung von Curricula, Assessmentinstrumenten oder Handlungsempfehlungen. Aufgrund der Flexibilität und geringen Standardisierung der Methode stehen Forschende vor der Herausforderung, zahlreiche methodische Entscheidungen vor und im Verlauf einer Delphi-Befragung zu treffen. Um ein strukturiertes und zielgerichtetes Vorgehen sicherzustellen, ist eine sorgfältige Planung vor Beginn einer Delphi-Studie unerlässlich.

Planung von Delphi-Studien: Die Planung von Delphi-Befragungen wird in diesem Beitrag in folgenden fünf Schritten beschrieben: 1. Eignung und Durchführbarkeit der Methode, 2. Forschungsfrage und beteiligte Personen, 3. Planung des Befragungsprozesses bis zur ersten Runde, 4. Auswertstrategien und Planung der Folgerunden, 5. Darstellung und Verbreitung der Ergebnisse. Jeder Schritt wird anhand zentraler Fragen strukturiert. Die jeweils wichtigsten Punkte werden in einer Checkliste zusammengefasst.

Schlussfolgerung: Dieser Leitfaden bietet Forschenden einen umfassenden Überblick über die methodischen Möglichkeiten und Grenzen von Delphi-Befragungen und weist auf mögliche Fallstricke hin. Er unterstützt bei der vorausschauenden Planung und hilft, fundierte Entscheidungen zu treffen. Langfristig kann dadurch die Qualität von Delphi-Studien in der medizinischen Ausbildungsforschung gesteigert und das Potenzial der Methode besser ausgeschöpft werden.

Schlüsselwörter: Delphi-Verfahren, Ausbildungsforschung, Curriculum, Forschungsdesign, Wissensgenerierung

Angelika Homberg¹

1. Medizinische Fakultät
Mannheim, Universität
Heidelberg, Geschäftsbereich
Studium und
Lehrentwicklung, Abteilung
Medizinische
Ausbildungsforschung,
Mannheim, Deutschland

1. Einleitung

Die medizinische Ausbildungsforschung umfasst ein weites Feld, welches auch die Entwicklung von konsensbasierten normativen Vorgaben und Empfehlungen einschließt [1]. In diesem Zusammenhang gewinnen Delphi-Befragungen zunehmend an Bedeutung, da sie den strukturierten Austausch zwischen unterschiedlichen Expert*innen ermöglichen und schließlich zu belastbaren und differenzierten Ergebnissen führen [2], [3]. Sie sind durch folgende Merkmale gekennzeichnet: Es erfolgt eine anonyme Befragung von Expert*innen durch standardisierte Fragebögen, oft mit offenen Fragen zur Erfassung von Argumenten, in mindestens zwei Runden. Die statistische Analyse beruht in der Regel auf deskriptiven Berechnungen. Ab der zweiten Runde erhalten die Expert*innen ein Feedback zu den Ergebnissen der Vorrunde und können so ihre Einschätzungen überdenken und gegebenenfalls revidieren. Auch der Fragebogen kann auf Basis der Rückmeldungen der Befragten angepasst werden [4], [5], [6].

Delphi-Befragungen können sowohl für qualitative als auch für quantitative Analysen genutzt, auf unterschiedliche Kontexte angepasst und somit vielseitig eingesetzt werden. Sie eignen sich insbesondere für die Bearbeitung komplexer Sachverhalte bei denen das Wissen unsicher oder interpretationsbedürftig ist [6]. Im Verlauf können unterschiedliche Perspektiven aufgegriffen und ein Konsens zu einer bestimmten Aufgabenstellung erarbeitet werden [7], [8].

Die Aufgabenstellungen in der medizinischen Ausbildungsforschung umfassen beispielsweise:

- die Überarbeitung oder Entwicklung von Curricula (z. B. [9], [10]),
- die Definition von Absolventenprofilen, Kompetenzen oder Lernzielen (z. B. [11], [12], [13]),
- die Überarbeitung oder Entwicklung von Items für Assessmenttools (z. B. [14], [15], [16]),
- die Entwicklung von outcome-basierten Indikatoren für die Beurteilung von Lernerfolgen (z. B. [17], [18]),
- das Sammeln und Gewichten von Ideen und Einflussfaktoren für die Weiterentwicklung innovativer Bereiche (z. B. [19], [20]),

- die Durchführung von Bedarfsanalysen (z. B. [21],[22]).

Delphi-Befragungen werden derzeit mit sehr heterogenen Konzepten und unterschiedlicher methodologischer Fundierung und Güte umgesetzt [23], [24]. Der Begriff *Modifikation* wird hierbei uneinheitlich verwendet [7], [25], [26]. Häufig werden zusätzlich face-to-face Meetings in den Befragungsprozess integriert, um durch den unmittelbaren Austausch die Chance einer Anwendung der Ergebnisse in der Praxis zu erhöhen [27], [28]. In diesem Falle wird die Anonymität der Teilnehmenden aufgehoben. Delphi-Befragungen erfordern aufgrund ihrer Komplexität eine sorgfältige Planung des gesamten Forschungsprozesses. Die Flexibilität der Methodik und fehlende Standardisierung erhöhen die Anforderungen an die Forschenden. Häder stellt in dem 2014 publizierten Standardwerk *Delphi-Befragungen* einige wichtige Aspekte mit Fokus auf die Bereiche Sozialwissenschaften und Zukunftsforschung detailliert dar [29]. Die derzeit existierenden Empfehlungen für die Durchführung von Delphi-Befragungen in den Gesundheitswissenschaften (z. B. [23], [25], [30], [31]) und in der medizinischen Ausbildungsforschung (z. B. [32], [33], [34]) wurden vor 2020 veröffentlicht. In den letzten Jahren hat sich die Methodik aufgrund technischer und erkenntnistheoretischer Entwicklungen weiter ausdifferenziert. Hierzu trug nicht zuletzt auch das DFG-geförderte Netzwerk DeWiss (Delphi-Verfahren in den Gesundheits- und Sozialwissenschaften) bei, das von 2020 bis 2023 neben konkreten Anwendungen auch grundlegende Fragen der Delphi-Methode thematisierte sowie eine internationale Delphi-Studie durchführte, um

eine umfassende Reporting Guideline für die Gesundheits- und Sozialwissenschaften zu entwickeln [35].

Das Ziel dieses *Gewusst wie* ist, Forschenden, die eine Delphi-Befragung im Bereich der medizinischen Ausbildungsforschung durchführen möchten, einen Überblick über die Möglichkeiten und Herausforderungen dieser Methode zu geben, damit potenzielle Fallstricke frühzeitig erkannt und valide, fundierte Ergebnisse erzielt werden können. Forschende werden in fünf Schritten anhand von Leitfragen durch den Prozess geführt. Am Ende werden alle wichtigen Punkte in einer Checkliste zusammengefasst (siehe Anhang 1). Diese kann auch erfahrenen Forschenden zur zielgerichteten und strukturierten Planung dienen.

2. Planung von Delphi-Befragungen

2.1. Schritt 1: Eignung und Durchführbarkeit der Methode

2.1.1. Ist die Delphi-Befragung die geeignete Methode?

Zunächst muss geklärt sein, welches spezifische Problem zu untersuchen ist. Ob die Delphi-Methode geeignet ist, hängt in erster Linie von den Rahmenbedingungen ab, die das Problem umgeben und nicht von der eigentlichen Fragestellung [6]. Aspekte, die für die Anwendung der Delphi-Methode sprechen, sind in Anlehnung an Linstone und Turoff [36] in Tabelle 1 aufgeführt.

Tabelle 1: Aspekte die zur Anwendung der Delphi-Methode führen (in Anlehnung an Linstone & Turoff 1975 [36])

Bereich	Aspekte
Untersuchungsgegenstand	Das Problem eignet sich nicht für präzise Analysetechniken, sondern profitiert von subjektiven Einschätzungen auf kollektiver Basis.
Beteiligte Personen	Die Personen, die zur Untersuchung des Problems beitragen sollen, konnten bislang nicht ausreichend miteinander kommunizieren.
	Die Personen repräsentieren möglicherweise unterschiedliche Hintergründe in Bezug auf Erfahrung oder Fachwissen.
	Es werden mehr Personen benötigt, als in einem persönlichen Austausch effektiv interagieren können.
Kommunikationsprozess	Zeit und Kosten machen häufige Gruppensitzungen undurchführbar.
	Die Effizienz von persönlichen Treffen kann durch einen ergänzenden Gruppenkommunikationsprozess gesteigert werden.
Gruppendynamik	Die Meinungsverschiedenheiten zwischen den einzelnen Teilnehmenden sind so schwerwiegend oder heikel, dass der Kommunikationsprozess moderiert werden muss.
	Die Anonymität der Teilnehmenden muss gewährleistet werden, damit diese ohne äußeren Bewertungsdruck antworten können.
	Bestehende Hierarchien verhindern, dass alle Personen gleichermaßen zum Prozess beitragen können.
	Die Heterogenität der Teilnehmenden muss ausgewogen abgebildet werden, um die Validität der Ergebnisse zu sichern, d.h. es muss vermieden werden, dass die Ergebnisse von der Anzahl der Teilnehmenden einer bestimmten Interessengruppe oder von der Stärke der Persönlichkeit einzelner („Mitläufereffekt“) dominiert werden.

Treffen wenige oder keine dieser Aspekte zu, sollte überlegt werden, ob andere Konsensmethoden wie eine Konsensuskonferenz oder die Nominal Group Technik [37] in Frage kommen oder eine Gruppendiskussion, eine Expert*innenbefragung oder ein Workshop die bessere Alternative wären.

2.1.2. Welche Ressourcen stehen zur Verfügung?

Die Komplexität der Befragung wirkt sich erheblich auf den Umfang der aufzubringenden Ressourcen aus. Bei guter Planung können Delphi-Befragungen auch von relativ kleinen Forschergruppen selbstständig und innerhalb weniger Monate durchgeführt werden. Auch die Durchführung im Rahmen einer Doktorarbeit ist realistisch [38]. Häufig unterschätzt werden die benötigten zeitlichen Ressourcen. Für jede Befragungsrunde sollten Zeitslots für die Nacherfassung von Non-Respondern eingeplant und Ferienzeiten sowie Zeiten mit hoher Arbeitsauslastung (z. B. Prüfungszeiten) berücksichtigt werden. Realistisch sind etwa acht bis zehn Wochen pro Runde. Zwischen den Befragungsrunden müssen innerhalb kurzer Zeit die Fragebögen ausgewertet, das Feedback vorbereitet, sowie der Fragebogen für die Folgerunde entwickelt und einem Prätest unterzogen werden, um die Motivation der Teilnehmenden aufrechtzuerhalten und Abbruchquoten zu reduzieren.

Finanzielle Ressourcen können für Incentives, Workshops (z. B. Raummieten und Reisekosten), professionell erstellte Begleitmaterialien oder Lizenzen für Befragungstools und Programme für die Auswertung der Daten benötigt [39] werden. Für einfache Delphi-Befragungen können übliche Surveytools verwendet werden, welche die Anonymität der Teilnehmenden und die Datensicherheit gewährleisten (z. B. [40]). Spezialisierte Tools lohnen sich bei aufwändigeren Verfahren, da sie Prozesse wie Verwaltung, Auswertung und Rückmeldung der Daten automatisieren sowie Vorlagen bereitstellen und Multi-Channel-Umfragen ermöglichen. Mit einzelnen Tools können Echtzeitberechnungen der Gruppenergebnisse im Rahmen eines Real-Time-Delphis durchgeführt werden [41].

2.1.3. Wann kommen Delphi-Varianten zum Einsatz?

In den letzten Jahrzehnten haben sich diverse methodische Varianten etabliert [8], welche einzelne typische Merkmale einer Delphi-Befragung außer Kraft setzen. So gibt es bei *Real-Time-Delphis* keine trennbaren Delphi-Runden [41]; beim *Gruppendelphi* wird die Anonymität der Expert*innen aufgehoben [42] und beim *Deliberativen Delphi* werden von Fachleuten geschulte Bürger*innen anstelle von Expert*innen eingebunden [43]. Der Einsatz eines *Real-Time Delphis* oder *Gruppendelphis* kann dazu beitragen, die Zeitdauer erheblich zu verkürzen. Im Bereich der medizinischen Ausbildungsforschung kommt vor allem das *Gruppendelphi* in Frage, um beispielsweise die synchrone Diskussion der Teilnehmenden

zur Konsensfindung eines komplexen Sachverhalts zu nutzen. Auch ein *Deliberatives Delphi* unter Einbindung von Patient*innen (z. B. [44]) oder Studierenden (z. B. [11], [45], [46]) ist für bestimmte Fragestellungen sinnvoll. Die Publikation von Niederberger und Deckert bietet einen guten Überblick über die Charakteristik der derzeit etablierten Varianten [8].

2.2. Schritt 2: Forschungsfrage und beteiligte Personen

2.2.1. Wer wird in die Planung und Steuerung des Prozesses eingebunden?

Um die mit einer Delphi-Befragung verbundenen Entscheidungen begründet und möglichst objektiv zu treffen, sollte eine Steuerungsgruppe zusammengestellt werden. In der Regel nehmen die Mitglieder der Steuerungsgruppe nicht selbst an der Delphi-Befragung teil. Eine Ausnahme bildet der partizipative Ansatz, bei welchem die Personen der Steuerungsgruppe gleichzeitig auch das Expert*innenpanel bilden können (z. B. [44]).

Generell kommen für die Steuerungsgruppe folgende Personen in Frage:

- Forschende, die die Delphi-Befragung durchführen.
- Forschende, die Erfahrung mit der Delphi-Methode haben.
- Forschende, die mit den Hintergründen des zu untersuchenden Problems vertraut sind.
- Entscheidungsträger, Stakeholder, etc., wie beispielsweise Studiendekane (strategischer Ansatz).
- Betroffene, wie beispielsweise Studierende oder Patient*innen (partizipativer Ansatz).

Durch den strategischen oder partizipativen Ansatz kann die Akzeptanz der Ergebnisse beziehungsweise deren Passung an konkrete Rahmenbedingungen erhöht werden.

Zur Bewältigung der vielfachen Aufgaben kann es sinnvoll sein, Hilfskräfte einzubinden [29]. Wichtig ist, die Verantwortungsbereiche und Aufgabenverteilung in der Steuerungsgruppe klar zu kommunizieren und die Planung sowie die im Verlauf getroffenen Entscheidungen konsequent zu protokollieren.

2.2.2. Wie wird die Forschungsfrage festgelegt?

Die erste Aufgabe der Steuerungsgruppe besteht darin, das Forschungsziel zu überprüfen und eine präzise Forschungsfrage zu formulieren. Dadurch wird sichergestellt, dass alle Mitglieder das gleiche Verständnis des Untersuchungsgegenstandes und der Zielsetzung haben und Missverständnisse vermieden werden.

Tabelle 2: Ausgewogene Zusammenstellung des Expert*innenpanels je nach Zielsetzung der Delphi-Befragung

Auswahlkriterien	Beispiel
Berufsgruppe	Interprofessionelle Studien: Ärzt*innen, Pflegekräfte und Therapeut*innen
Region bzw. Nation	Internationale Studien: Teilnehmenden aus strukturstarken und -schwachen Ländern
Hierarchieebene	Entwicklung von Kriterien für den klinischen Unterricht: Oberärzt*innen und Assistenzärzt*innen
Erfahrungsbereich	Entwicklung von Absolvent*innenprofilen: Expert*innen aus dem ambulanten und stationären Bereich
Perspektive	Entwicklung von longitudinalen Curricula für klinische Fertigkeiten: Studierende aus verschiedenen Ausbildungsphasen, Dozierende, Patient*innen
Soziodemografische Daten, wie Geschlecht, Alter und institutionelle Zugehörigkeit	Entwicklung von gendergerechten Lernmaterialien: weibliche und männliche Studierende

Hierbei sollte Folgendes überlegt werden:

- ob Fachbegriffe von allen Beteiligten gleichermaßen genutzt und verstanden werden,
- ob die Ergebnisse nur für eine bestimmte Region, national oder international anwendbar sein sollen,
- unter welchen Bedingungen die Ergebnisse genutzt werden sollen (z. B. für Schulen, Universitäten oder Weiterbildungseinrichtungen),
- welches methodische Ziel insgesamt erreicht werden soll (z. B. Konsensfindung, Ideenaggregation, Erhebung von Expert*innenmeinungen).

In der medizinischen Ausbildungsforschung sind Delphi-Befragungen zur Konsensfindung vorherrschend. Die Steuerungsgruppe sollte prüfen, ob andere Ziele, wie beispielsweise die Ideenaggregation oder die Erhebung der Expert*innenmeinungen im Vordergrund stehen. Über die jeweilige Ausdifferenzierung liefert Häder einen guten Überblick [29].

2.2.3. Welche Personengruppen werden befragt?

Die Befragten bei Delphi-Studien werden in der Regel Expert*innen genannt, welche als informierte Individuen oder Spezialist*innen über Wissen in einen bestimmten Bereich verfügen [25]. Sie bringen als reflektierte Praktiker*innen ihre lebensweltlichen Erfahrungen in den Delphi-Prozess ein [27]. Die Zusammenstellung des Expert*innenpanels ist wesentlich von der methodischen Zielsetzung der Delphi-Befragung abhängig [29].

Bei Konsensdelphi-Studien sollte die Vielfalt der Grundgesamtheit derer abgebildet werden, welche über einschlägige Expertise zum Forschungsthema verfügen. Alle hierfür relevanten Gruppen sollten so ausgewogen wie möglich in das Delphi integriert werden [47], [48] (siehe Tabelle 2) und jeweils durch mindestens zwei, besser drei Personen vertreten sein. Dadurch können unterschiedliche Standpunkte innerhalb der jeweiligen Gruppe erfasst werden und ein Drop-out einzelner Personen hat weniger Einfluss auf das Gesamtergebnis. Eine Nachrekrutierung von Expert*innen ist grundsätzlich möglich.

Sie sollte insbesondere dann in Erwägung gezogen werden, wenn die Gefahr besteht, dass der Drop-out zu einer Verzerrung der Ergebnisse führt (wenn überproportional Expert*innen aus einer Gruppe ausscheiden) oder dass sich die statistische Aussagekraft durch die Reduzierung der Stichprobengröße reduziert (bei kleinen Stichproben). Um die Vergleichbarkeit, Konsistenz und Integrität der Studie zu wahren, sollten die neuen Teilnehmenden eine vergleichbare Expertise aufweisen und mit den bisherigen Ergebnissen und dem Verlauf vertraut gemacht werden. Die Forschenden müssen die Nachrekrutierung offenlegen und begründen.

Eine große Schwierigkeit besteht darin, zu klären, was Expert*innen tatsächlich zu Expert*innen macht [49]. Diese Frage sollte in Bezug auf die Zielsetzung der Delphi-Befragung in der Steuerungsgruppe diskutiert werden, da von der Zusammensetzung des Expert*innenpanels maßgeblich die Reliabilität und Validität der Studie beeinflusst werden.

2.2.4. Wie wird die Stichprobe ermittelt?

Da bei Delphi-Befragungen in der Regel keine Totalerhebungen möglich sind, muss eine Stichprobe ermittelt werden. Für den Erfolg der Befragung sind hierzu vor allem zwei Faktoren relevant: die individuelle Expertise der Befragten in Bezug auf die Aufgabenstellung sowie ihr persönliches Interesse am Thema. Durch die gezielte Auswahl kann der Drop-out erheblich gesenkt werden [50]. Um systematisch vorzugehen, können folgende Kriterien herangezogen werden: Anzahl der veröffentlichten Artikel zu dem betreffenden Themenbereich oder Publikationskennzahl, Mitarbeit in Fachgremien, berufliche Position oder Aktivitäten, aktive Beteiligung in einschlägigen Netzwerken und Foren. Möglich ist auch, dass Fakultäten oder Hochschulen angefragt werden, Expert*innen zu einem bestimmten Themenbereich zu benennen. Studierende können beispielsweise über die Fachschaft rekrutiert werden, Patient*innen über Selbsthilfegruppen. Persönliche Kontakte und Zugehörigkeiten zu einschlägigen Netzwerken der Mitglieder der Steuerungsgruppe sind hilfreich, um motivierte Personen ausfindig zu machen.

2.2.5. Wie viele Expert*innen werden eingeschlossen?

Es gibt keine einheitlichen Empfehlungen zur Anzahl der einzuschließenden Expert*innen. Laut Turoff [51] ist eine Anzahl von zehn bis vierzig Personen günstig. Trevelyan et al. sprechen von acht bis fünfzehn Teilnehmenden bei homogenen Panels im gesundheitswissenschaftlichen Bereich [30]. Bei Delphi-Befragungen in der medizinischen Ausbildungsforschung zeigen sich Panels von acht (z. B. [38]) bis über hundert (z. B. [52]).

Der Zeitaufwand erhöht sich mit der Teilnehmerzahl, insbesondere wenn die Befragung einen großen Anteil qualitativer Fragen beinhaltet und Arbeitsschritte nicht automatisiert ablaufen.

Ein kleines Expert*innenpanel reicht unter folgenden Voraussetzungen aus:

- Es müssen nur wenige Themen, Thesen und Fragen bearbeitet werden.
- Es müssen nur wenige verschiedene Expertisen, Standpunkte und Perspektiven berücksichtigt werden.
- Es gibt eine kleine Grundgesamtheit, die abgebildet werden kann.
- Die Delphi-Befragung lässt sich in einem kurzen Zeitraum durchführen.
- Die Motivation der Teilnehmenden, und dadurch auch der erwartete Rücklauf, wird eher hoch eingeschätzt.
- Die erwartete Bereitschaft der Teilnehmenden, Argumente und Ideen einzubringen wird hoch eingeschätzt.
- Es besteht die Möglichkeit, ausfallende Expert*innen nach der ersten Runde nachzurekrutieren.

2.3. Schritt 3: Planung des Befragungsprozesses bis zur ersten Runde

2.3.1. Wie werden die Items für den ersten Fragebogen entwickelt?

Für den Bereich der Ausbildungsforschung beschreiben Garavalia et al. [53] in Anlehnung an Linstone [36] folgendes Vorgehen als klassisches Delphi in drei Runden:

1. *Runde* Sammeln von Items anhand einer offenen Fragestellung,
2. *Runde* Bewertung der gesammelten Items anhand von Ratingskalen und zusätzliches Einholen von Argumenten mittels offener Fragen,
3. *Runde* Erneute Bewertung der Items auf Grundlage der rückgespiegelten Kommentare und Argumente [53].

Dieses Vorgehen wird in der medizinischen Ausbildungsforschung nur vereinzelt gewählt (z. B. [11], [12]). Häufiger werden Items vorab formuliert. Die teilnehmenden Expert*innen werden gebeten, diese zu bewerten und weitere Items vorzuschlagen. Der Delphi-Prozess kann dadurch beschleunigt werden. Zudem erhalten die Ex-

pert*innen eine Vorstellung davon, was von ihnen erwartet wird und generieren eher vergleichbare neue Inhalte. Der Workload für die Befragten und die Wahrscheinlichkeit, vom Thema abzuweichen, werden reduziert.

Die Generierung der Items für die erste Befragungsrunde richtet sich nach den verfügbaren Ressourcen [29]. Bei bestimmten Fragestellungen können Aussagen aus von Fachgesellschaften oder -gremien erstellten Listen übernommen werden. Häufig existieren einschlägige Publikationen, so dass die Items literaturbasiert durch die Steuerungsgruppe zusammengestellt werden können. Wenn diese Voraussetzungen nicht erfüllt sind, können die Items auch direkt durch die Steuerungsgruppe entwickelt werden. In diesem Fall ist es ratsam, zusätzliche Expert*innen, etwa aus Fachausschüssen und Gremien, in einem vor der Delphi-Befragung stattfindenden Expert*innenmeeting oder durch eine Vorabbefragung einzubeziehen.

2.3.2. Wie wird die Anzahl der Runden festgelegt?

Die Anzahl der Runden sollte zu Beginn der Studie feststehen und unabhängig davon sein, ob für alle Items ein Konsens erreicht wurde. Wenn das Erreichen des Konsenses als Stopp-Kriterium festgelegt wird, bedeutet dies, dass die Befragung solange fortgeführt wird, bis alle Items eindeutig zugeordnet werden konnten. Der Aufwand ist somit weder für die Steuerungsgruppe noch für die Teilnehmenden kalkulierbar.

Für die Festlegung der Rundenzahl ist entscheidend, wie gut ein Thema bereits erforscht ist, wie heterogen das Expert*innenpanel ist, wie der Konsens definiert wird, welches Ergebnis erzielt werden soll und welche Ressourcen zur Verfügung stehen. Bei einer Delphi-Studie von Ellaway et al. [10] sollte beispielsweise ein Curriculum für das Medizinstudium zur Versorgung von Trans- und geschlechtsdiversen Patient*innen entwickelt werden. Die Komplexität erforderte vier Runden. Häufig lassen sich aber auch in zwei Runden schon gute Ergebnisse erzielen (z. B. [54], [55], [56]).

2.3.3. Wie wird Konsens gemessen?

Konsens bei Delphi-Befragungen bedeutet, dass eine hohe Übereinstimmung zwischen den befragten Expert*innen zu einer Aussage erreicht wird und nicht, dass alle Expert*innen dieselbe Meinung teilen. In den meisten Publikationen wird ein Konsens bei etwa 70-85% festgelegt [57]. Die Übereinstimmung wird in der Regel über deskriptive statistische Maße gemessen und ist abhängig vom Skalenniveau [58]. Bei einem nominalen Skalenniveau kann die relative Häufigkeit der Nennungen bestimmt werden. In der Regel wird der Grad der Zustimmung auf einer ordinalen Skala erfasst. Auch hier kann der Anteil der Befragten bestimmt werden, welcher einer Aussage ab einer bestimmten Merkmalsausprägung zustimmt oder diese ablehnt. Häufiger werden Median und Interquartilsabstand (IQR) herangezogen, da dadurch

Tabelle 3: Aufbau des Fragebogens für die erste Delphi-Runde

Kapitel	Inhalt	Anmerkung
Präambel	• Ziel und Kontext der Befragung • Verwendung zentraler Begriffe • Aufbau des Fragebogens • Bearbeitungszeit • Prozess der Ergebnisaufbereitung und Rückspiegelung • Befragungsprozess insgesamt	Wenn die Befragten umfangreiche Hintergrundinformationen benötigen oder viele Fachbegriffe erläutert werden müssen, können diese als Begleitmaterial oder Glossar per Mail verschickt oder direkt im Fragebogen verlinkt werden (Mouse over).
Erfassung der demographischen Daten und Interessenkonflikte	• Merkmale, um die Expertise, Perspektiven und mögliche Interessenskonflikte des Expert*innenpanels transparent darstellen zu können	Wenn nur vollständig ausgefüllte Fragebögen ausgewertet werden, kann dieser Block auch an das Ende der Befragung gestellt werden.
Hauptteil geschlossene Fragen	• Darstellung der einzelnen Aussagen bzw. Items als Thesen • Bewertung auf einer 3 bis 9-stufigen Ratingskala plus der Option „keine Aussage“	Die Auswertmöglichkeiten für die Forschenden, aber auch der kognitive Aufwand für die Befragten erhöhen sich mit der Anzahl der Stufen. Eine mittlere Option wirkt der Verzerrung der Ergebnisse entgegen.
Hauptteil offene Fragen	Offene Fragen in Kombination mit Thesen: • Erfassung von Argumenten für eine Beurteilung (<i>argumentative Fragen</i>) • Erfassung weiterer Items (<i>Fragen zur Extension</i>) Alleinstehende Fragen • Sammeln von Ideen zu einem bestimmten Sachverhalt (<i>Fragen zur Exploration</i>)	Vorteil offener Fragen: Die Befragten können in ihren eigenen Worten antworten, Forschende erhalten wertvolle Einsichten in deren Beweggründe. Nachteil offener Fragen: Erfordern einen höheren kognitiven und zeitlichen Aufwand für die Befragten, es besteht die Gefahr, dass die Bereitschaft zu antworten geringer ausfällt.

Mehrheiten abgebildet werden können [59]. Bei einer 5-stufigen Ratingskala kann beispielsweise ein IQR <1 als Konsens definiert werden [58]. Die Verwendung von Mittelwert und Standardabweichung ist bei ordinalen Skalen zu vermeiden. Konsens kann auch über Rankingfragen ermittelt werden. Hier muss aber berücksichtigt werden, dass sie trotz des höheren kognitiven Aufwands für die Befragten in der Regel keinen Mehrwert gegenüber Ratingfragen vorweisen können [60].

2.3.4. Wie ist der erste Fragebogen aufgebaut?

Der Fragebogen für die erste Delphi-Runde sollte eine Präambel enthalten, demographische Daten erheben und im zentralen Teil offene und geschlossene Fragen kombinieren. Die Skalenstufen und Optionen für die geschlossenen Fragen müssen begründet ausgewählt werden [28], [59], [61] (siehe Tabelle 3).

Eine klare Struktur des ersten Fragebogens und eine überschaubare Menge erfasster Daten tragen wesentlich dazu bei, die Folgerunden und das Feedback nicht zu überladen und angemessen gestalten zu können [30].

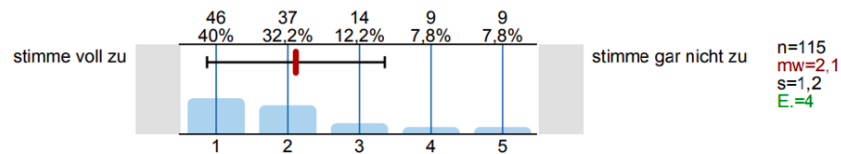
Belton et al. empfehlen, dass die Bearbeitungszeit einer Delphi-Runde nicht mehr als 30 Minuten betragen und in der ersten Runde nur wenige offene Fragen integriert werden sollten [62].

2.3.5. Wie wird mit den Expert*innen kommuniziert?

Wichtig ist, die teilnehmenden Expert*innen als Partner*innen zu behandeln und alle wichtigen Schritte zu erklären. Dies beginnt damit, die Expertise explizit zu benennen, Befragte also nicht als *Expert*innen* zu bezeichnen, ohne den Begriff erläutert zu haben [30]. Die Anfrage zur Teilnahme sollte persönlich gestaltet werden. Wird begründet, warum gerade die individuelle Expertise wichtig ist, wächst die Bereitschaft. Um die Motivation aufrecht zu erhalten, sind soziale Belohnungen häufig sinnvoller als finanzielle [62], z. B. wenn die Teilnehmenden erkennen, dass sie durch ihre Antworten tatsächlich einen Einfluss auf das Ergebnis der Delphi-Befragung haben oder Hintergrundinformationen und Zwischenergebnisse exklusive bereitgestellt werden. Weiterhin ist

Item 4.3: Open-Book-Prüfungen sollten künftig häufiger als Leistungsnachweis eingesetzt werden

Statistik



Argumente

Zustimmung	Ablehnung
- Kompetenz-orientiertes Prüfen wird gefördert - Fragengestaltung ist flexibler	- Korrekturaufwand zu hoch - Bewertungen schwer zu objektivieren

Abbildung 1: Beispiel für die Gestaltung des Feedbacks bei Delphi-Befragungen

die angemessene Sprache im Fragebogen und bei den Begleitmaterialien maßgeblich für den Erfolg. Insbesondere bei heterogenen Expert*innenpanels, wie bspw. bei internationalen Studien oder partizipativen Delphi-Befragungen, sollten im Pretest Vertreter*innen aus den einzelnen Untergruppen des Expert*innenpanels eingebunden werden, um die Verständlichkeit für alle Teilnehmenden zu gewährleisten [63].

2.4. Schritt 4: Auswertstrategien und Planung der Folgerunden

2.4.1. Welche Strategien werden für die Auswertung verwendet?

Die Auswertungsstrategien sind abhängig vom Datenmaterial. Quantitative Daten werden in der Regel deskriptiv statistisch ausgewertet. Für die Konsensmessung wird jedes Item einzeln betrachtet und auf Grundlage der festgelegten Konsenskriterien in die Gruppen: *Konsens über Zustimmung*, *Dissens* oder *Konsens über Ablehnung* eingeordnet.

Antworten auf offene Fragen werden in Abhängigkeit ihrer Funktion ausgewertet [64] (siehe Tabelle 3). Fragen zur Exploration können anhand der quantitativen Inhaltsanalyse [65] oder Thematischen Analyse [66] kategorisiert und zusammengefasst werden. Fragen zur Extension können anhand der quantitativen Inhaltsanalyse ausgewertet werden [67]. Eine besondere Rolle spielt die Auswertung von qualitativ erfassten Argumenten, bei denen die Befragten ihre quantitative Bewertung begründen. Hierfür wurde explizit die fünf-stufige *Argumentenbasierte Qualitative Analyse* AQUA entwickelt, welche die Möglichkeit bietet, die qualitativen Daten strukturiert auszuwerten und gemeinsam mit den quantitativen Daten für das Feedback aufzubereiten [68].

2.4.2. Wie wird das Feedback übermittelt?

Die Teilnehmenden sollten nach jeder Delphi-Runde einen objektiven Feedbackbericht über die aggregierten Gruppenergebnisse erhalten. Hierbei sollte anonym dargestellt werden, wie sich die Gruppe der Teilnehmenden zusam-

mensetzte. Quantitative Daten können anhand der Lage- und Streumaße grafisch visualisiert werden, z. B. als Balken- oder Säulendiagramm. Individualisierte Feedbacks, bei welchen die jeweils eigene Bewertung im Vergleich zu den Gruppenwerten rückgespielt werden, sind in der Regel aufwändig und werden im Hinblick auf den Nutzen kritisch diskutiert, da hierdurch ein Konformitätsdruck und höhere kognitive Belastung entstehen [69]. Die qualitativen Ergebnisse sollten inhaltlich gut strukturiert und nahe am ursprünglichen Wortlaut zurückgespielt werden, so dass die Expert*innen ihre eigenen Beiträge erkennen können. Ziel ist es, dass die Teilnehmenden auf Basis der Gruppenantworten ihre Meinung überdenken und möglicherweise anpassen können. Abbildung 1 zeigt ein Beispiel, wie die quantitativen Daten mit dem individuellen Feedback und den Argumenten gebündelt dargestellt werden können.

Der Feedbackbericht kann mit der Einladung zur Folgerunde an die Befragten als separates Dokument verschickt oder in den Folgefragebogen integriert werden.

2.4.3. Wie wird mit dem ermittelten Konsens umgegangen?

Items, bei denen ein Konsens über eine Zustimmung oder Ablehnung ermittelt wurde, werden im Zwischen- und Abschlussbericht dargestellt, aber nicht erneut in der Folgerunde zur Bewertung vorgelegt. Zwar könnten bei einer erneuten Abstimmung die Reproduzierbarkeit und damit die Stabilität der Bewertungen ermittelt werden, dieser Vorteil wiegt aber in der Regel nicht die zusätzliche Belastung für die Befragten auf.

Items, über die kein Konsens für eine Zustimmung oder Ablehnung erreicht wurde, werden analysiert [70] und entsprechend der Analyseergebnisse terminiert, modifiziert oder unverändert in Kombination mit den Argumenten erneut zur Bewertung vorgelegt [36] (siehe Abbildung 2).

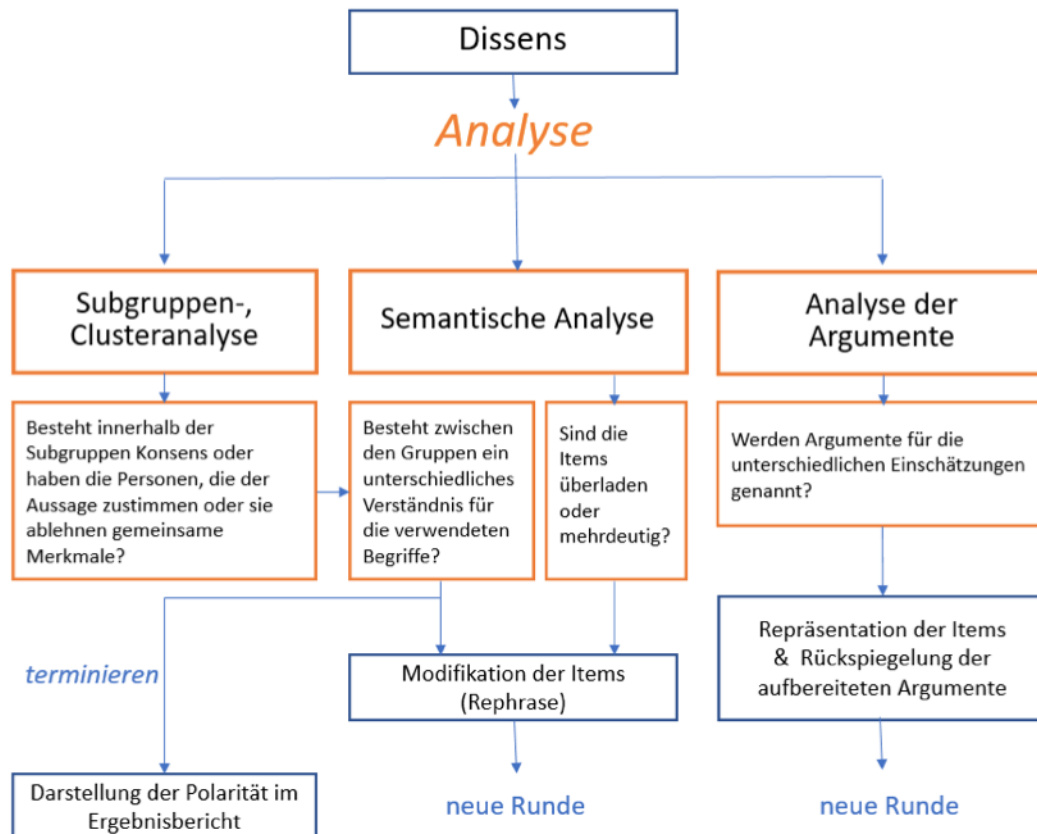


Abbildung 2: Analysetechniken für im Dissens verbliebene Items bei Delphi-Befragungen

2.4.4. Wie werden die Fragebogen der Folgerunden aufgebaut?

Fragebögen der Folgerunden greifen die Struktur des ersten Fragebogens auf und enthalten in der Regel neben den identischen und modifizierten Items aus der Vorrunde auch Items, welche aus den Antworten auf explorative Fragen neu generiert wurden. Weiterhin sollte die Möglichkeit gegeben werden, dass Expert*innen ihre Urteile begründen können. Offene Fragen zur Exploration sollten im Verlauf der Delphi-Befragung nicht mehr oder nur in einem begrenzten Umfang angeboten werden.

2.5. Schritt 5: Darstellung und Verbreitung der Ergebnisse

2.5.1. Was ist mit Items, die im Dissens bleiben?

Selbst nach vielen Runden kann es sein, dass zu einzelnen Items kein Konsens gefunden wurde. Das Nichterreichen von Konsens muss aufgegriffen und analysiert werden.

Es könnte beispielsweise sein, dass die Expert*innen

- generell eine unterschiedliche Priorisierung von Werten bzw. Zielen haben,
- Teilaspekte unterschiedlich gewichten oder berücksichtigen oder

- Interessenkonflikte haben und deshalb nicht von ihren Positionen abweichen.

Ursachen könnten auch in der methodischen Umsetzung liegen, z. B.

- dass das Themenfeld im Vorfeld nicht entsprechend eingegrenzt wurde,
- Argumente im Feedback nicht ausreichend kommuniziert oder berücksichtigt wurden,
- die Anzahl der Runden nicht ausreichend war oder
- die Items auch im Verlauf nicht präzise formuliert wurden.

Im Dissens gebliebene Aussagen können innerhalb der Steuerungsgruppe, in einem Abschlussmeeting mit den teilnehmenden Expert*innen oder unter Hinzuziehung weiterer Expert*innen oder Betroffenen weiter diskutiert werden. Unabhängig davon sollten dissente Urteile immer im Ergebnisbericht dargestellt werden.

2.5.2. Wie werden die Ergebnisse publiziert?

Um die Qualität und Transparenz von Berichterstattungen zu erhöhen, sollte eine entsprechende Reporting-Guideline verwendet werden. DelphiSTAR (Delphi studies in social and health sciences - Recommendations for an interdisciplinary standardized reporting) wurde speziell für Publikationen in den Gesundheits- und Sozialwissenschaften entwickelt [35]. Möglich ist auch die Orientierung an ACCORD (ACcurate CONsensus Reporting Document), welche für die Biomedizin entwickelt wurde [35], oder

CREDES (Guidance on Conducting and REporting DELphi Studies) für den Palliativbereich [31].

Die teilnehmenden Expert*innen bringen ihre Erfahrungen und erhebliche zeitliche Ressourcen in den Delphi-Prozess ein. Dies sollte in der Danksagung gewürdigt werden, wenn die entsprechende Einwilligung vorliegt, auch namentlich. Im entsprechenden Anhang kann die Qualifikation der Expert*innen dargestellt werden, um die in den Delphi-Prozess eingebrachte Expertise zu belegen oder Interessenkonflikte offenzulegen.

2.5.3. Wie werden die Ergebnisse verbreitet und implementiert?

Die Forschenden sollten sich schon vor Beginn der Studie damit beschäftigen, ob Maßnahmen zur Verbreitung und Implementierung der Ergebnisse notwendig sind. In vielen Fällen müssen die Ergebnisse noch weiter aufbereitet werden, um sie in die Praxis zu überführen. Beispielsweise können Zusammenfassungen formuliert und an die Zielgruppen verbreitet werden. Auch hierzu kann ein Abschlussworkshop hilfreich sein (z. B. [71], [72]). Je nach Outcome kann es auch sinnvoll sein, die Umsetzbarkeit der Ergebnisse in der Praxis zu evaluieren und weitere Möglichkeiten der Anwendung und Implementierung zu überprüfen.

4. Diskussion

Dieser Beitrag liefert eine komprimierte Anleitung zur Durchführung von Delphi-Befragung in der medizinischen Ausbildungsforschung. Wesentliche Arbeitsschritte werden benannt, übliche Vorgehensweisen und Alternativen beschrieben und punktuell auf Fallstricke hingewiesen. Es wird vorausgesetzt, dass Forschende mit den Grundlagen der empirischen Sozialforschung vertraut sind, beispielsweise mit Fragebogenkonstruktion, Operationalisierung theoretischer Begriffe oder Datenanalyse und Interpretation. Wenn Forschende Delphi-Varianten anwenden möchten, ist die Auseinandersetzung mit vertiefender Literatur unumgänglich.

Es ist zu bedenken, dass methodische Entscheidungen häufig weitreichende Konsequenzen haben. Beispielsweise wirkt sich die Zusammenstellung des Expert*innenpanels [29], die Wahl der Bewertungsskala und die Festlegung des Konsenses [59] erheblich auf die Ergebnisse aus. Die statistische Darstellung der Ergebnisse kann dazu verleiten, sie als objektive Messung fehlzuinterpretieren und die Aussagekraft zu überschätzen. In der medizinischen Ausbildungsforschung sind sie eher als Ergebnis eines situativen Aushandlungsprozesses zu verstehen, als Verständigung darüber, wie etwas sein sollte oder vollzogen werden könnte. Eine große Herausforderung besteht darin, dass viele Punkte in Bezug auf Delphi-Befragungen noch nicht abschließend geklärt sind, beispielsweise fehlt es an Sekundäranalysen, experimentellen Studien oder Methodentests, um fundierte Aussagen über den Einfluss methodischer und statistischer

Vorgehensweisen auf das Urteilsverhalten und damit auf die Konsensfindung machen zu können [3], [58]. Dennoch sind sie das Mittel der Wahl, um Fachwissen und die lebensweltliche Erfahrung von Expert*innen aus Wissenschaft und Praxis systematisch zusammenzuführen und dadurch mehr Planungs- und Handlungssicherheit zu erlangen. Dieses *Gewusst wie* kann dazu beitragen, die Qualität von Delphi-Studien in der medizinischen Ausbildungsforschung zu fördern. Die hier genannten Schritte und Empfehlungen können weitgehend auf andere Delphi-Studien im gesundheits- und bildungswissenschaftlichen Bereich übertragen werden.

ORCID der Autorin

Angelika Homberg: [0000-0001-5585-1126]

Interessenkonflikt

Die Autorin erklärt, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Artikel hat.

Anhänge

Verfügbar unter <https://doi.org/10.3205/zma001800>

1. Anhang_1.pdf (95 KB)
Checkliste für die Durchführung von Delphi-Studien in der medizinischen Ausbildungsforschung

Literatur

1. Schüttpeitz-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/zma001804
2. Foth T, Efstathiou N, Vanderspank-Wright B, Uffholz LA, Dutthorn N, Zimansky M, Humphrey-Murto S. The use of Delphi and Nominal Group Technique in nursing education: A review. *Int J Nurs Stud.* 2016;60:112-120. DOI: 10.1016/j.ijnurstu.2016.04.015
3. Humphrey-Murto S, Varpio L, Wood TJ, Gonsalves C, Uffholz LA, Mascioli K, Wang C, Foth T. The Use of the Delphi and Other Consensus Group Methods in Medical Education Research: A Review. *Acad Med.* 2017;92(10):1491-1498. DOI: 10.1097/acm.0000000000001812
4. Webler T, Levine D, Rakel H, Renn O. A novel approach to reducing uncertainty: The group Delphi. *Technol Forecast Soc Change.* 1991;39(3):253-263. doi: 10.1016/0040-1625(91)90040-M
5. Niederberger M, Renn O, editors. Delphi-Verfahren in den Sozial- und Gesundheitswissenschaften: Konzept, Varianten und Anwendungsbeispiele. Wiesbaden: Springer VS; 2019. DOI: 10.1007/978-3-658-21657-3
6. Linstone HA, Turoff M. The Delphi Method: Techniques and Applications. Boston: Addison-Wesley Publishing Company; 2002.

7. Homberg A, Schüttelz-Brauns K. Delphi-Befragungen in der Ausbildungsforschung – Anwendungsbereiche und Umsetzung. In: Gemeinsame Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA) und des Arbeitskreises zur Weiterentwicklung der Lehre in der Zahnmedizin (AKWLZ). Halle (Saale), 15.-17.09.2022. Düsseldorf: German Medical Science GMS Publishing House; 2022. DocV-10-04. DOI: 10.3205/22gma062
8. Niederberger M, Deckert S. Das Delphi-Verfahren: Methodik, Varianten und Anwendungsbeispiele [The Delphi technique: Methodology, variants and usage examples]. Z Evid Fortbild Qual Gesundhwes. 2022;174:11-19. DOI: 10.1016/j.zefq.2022.08.007
9. Gordon M, Baker P, Catchpole K, Darbyshire D, Schocken D. Devising a consensus definition and framework for non-technical skills in healthcare to support educational design: A modified Delphi study. Med Teach. 2015;37(6):572-577. DOI: 10.3109/0142159X.2014.959910
10. Ellaway RH, Thompson NL, Temple-Oberle C, Pacaud D, Frecker H, Jablonski TJ, Demers J, Mattatall F, Raiche J, Hull A, Jalil R. An undergraduate medical curriculum framework for providing care to transgender and gender diverse patients: A modified Delphi study. Perspect Med Educ. 2022;11(1):36-44. DOI: 10.1007/s40037-021-00692-7
11. Berger-Estilita J, Nabecker S, Greif R. A Delphi consensus study for teaching "Basic Trauma Management" to third-year medical students. Scand J Trauma Resusc Emerg Med. 2019;27(1):91. DOI: 10.1186/s13049-019-0675-6
12. Almeland SK, Lindford A, Berg JO, Hansson E. A core undergraduate curriculum in plastic surgery - a Delphi consensus study in Scandinavia. J Plast Surg Hand Surg. 2018;52(2):97-105. DOI: 10.1080/2000656X.2017.1343190
13. Behrend R, Herinek D, Kienle R, Arnold F, Peters H. Entwicklung interprofessioneller Ausbildungsziele für die Gesundheitsberufe an der Charité – Universitätsmedizin Berlin – Eine Delphi-Studie [Development of Interprofessional Learning Outcomes for Health Professions at Charite - Universitätsmedizin Berlin - A Delphi-Study]. Gesundheitswesen. 2022;84(6):532-538. DOI: 10.1055/a-1341-1368
14. Al Asmri M, Haque MS, Parle J. A Modified Medical Education Research Study Quality Instrument (MMERSQI) developed by Delphi consensus. BMC Med Educ. 2023;23(1):63. DOI: 10.1186/s12909-023-04033-6
15. Fluit C, Bolhuis S, Grol R, Ham M, Feskens R, Laan R, Wensing M. Evaluation and feedback for effective clinical teaching in postgraduate medical education: Validation of an assessment instrument incorporating the CanMEDS roles. Med Teach. 2012;34(11):893-901. DOI: 10.3109/0142159x.2012.699114
16. Kahr Rasmussen N, Nayahangan LJ, Carlsen J, Ekberg O, Brabrand K, Albrecht-Beste E, Bachmann Nielsen M, Konge L. Evaluation of competence in ultrasound-guided procedures-a generic assessment tool developed through the Delphi method. Eur Radiol. 2021;31(6):4203-4211. DOI: 10.1007/s00330-020-07280-z
17. Rau T, Fegert JM, Kornmann M, Muche R, Gulich M, Barth H, Kornmann M, Fegert JM, Öchsner W. Erarbeitung von Empfehlungen zu universitären Prüfungen an der Medizinischen Fakultät Ulm. GMS Z Med Ausbild. 2010;27(1):Doc07. DOI: 10.3205/zma000644
18. de Leeuw RA, Walsh K, Westerman M, Scheele F. Consensus on Quality Indicators of Postgraduate Medical E-Learning: Delphi Study. JMIR Med Educ. 2018;4(1):e13. DOI: 10.2196/mededu.9365
19. Homberg A, Stock-Schröer B. Interprofessional Education on Complementary and Integrative Medicine. Clin Teach. 2021;18(2):152-157. DOI: 10.1111/tct.13280
20. Teo YH, Peh TY, Abdurrahman A, Lee AS, Chiam M, Fong W, Wijaya L, Krishna LK. A modified Delphi approach to enhance nurturing of professionalism in postgraduate medical education in Singapore. Singapore Med J. 2024;65(6):313-325. DOI: 10.11622/smedj.2021224
21. Thim S, Nayahangan LJ, Paltved C, Jensen RD, Konge L, Hertel NT, Balslev T. Identifying and prioritising technical procedures for simulation-based curriculum in paediatrics: a Delphi-based general needs assessment. BMJ Paediatr Open. 2020;4(1):e000697. DOI: 10.1136/bmjpo-2020-000697
22. Bessmann EL, Østergaard HT, Nielsen BU, Russell L, Paltved C, Østergaard D, Konge L, Nayahangan LJ. Consensus on technical procedures for simulation-based training in anaesthesiology: A Delphi-based general needs assessment. Acta Anaesthesiol Scand. 2019;63(6):720-729. DOI: 10.1111/aas.13344
23. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One. 2011;6(6):e20476. DOI: 10.1371/journal.pone.0020476
24. Schifano J, Homberg A, Sonnberger M, Niederberger M. Reporting guidelines for Delphi techniques in health sciences: A methodological review. Z Evid Fortbild Qual Gesundhwes. 2022;172:1-11. DOI: 10.1016/j.zefq.2022.04.025
25. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. Int J Nurs Stud. 2001;38(2):195-200. DOI: 10.1016/S0020-7489(00)00044-4
26. Schifano J, Niederberger M. How Delphi studies in the health sciences find consensus: a scoping review. Syst Rev. 2025;14(1):14. DOI: 10.1186/s13643-024-02738-3
27. Homberg A, Hirt J, Niederberger N, Häder M, Vollmar H, Cuhls K, Reiber K, Mohr J. Kann Delphi nur Zukunftsforschung? – Ein Blick in die Gesundheitswissenschaften. Z Zukunftsforsch. 2024;12(1). Zugänglich unter/available from: <https://www.zeitschrift-zukunftsforschung.de/zukunftsforschung/article/view/7>
28. Shang Z. Use of Delphi in health sciences research: A narrative review. Medicine (Baltimore). 2023;102(7):e32829. DOI: 10.1097/MD.00000000000032829
29. Häder M. Delphi-Befragungen: Ein Arbeitsbuch. 3. Auflage. Wiesbaden: Springer VS; 2014. DOI: 10.1007/978-3-658-01928-0
30. Trevelyan EG, Robinson PN. Delphi methodology in health research: how to do it? Eur J Integr Med. 2015;7(4):423-428. DOI: 10.1016/j.eujim.2015.07.002
31. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on Conducting and Reporting DELphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. Palliat Med. 2017;31(8):684-706. DOI: 10.1177/0269216317690685
32. Waggoner J, Carline JD, Durning SJ. Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research. Acad Med. 2016;91(5):663-668. DOI: 10.1097/ACM.0000000000001092
33. Donohoe H, Stellefson M, Tennant B. Advantages and Limitations of the e-Delphi Technique. Am J Health Educ. 2012;43(1):38-46. DOI: 10.1080/19325037.2012.10599216
34. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and Nominal Group in medical education research. Med Teach. 2017;39(1):14-19. DOI: 10.1080/0142159x.2017.1245856

35. Niederberger M, Schifano J, Deckert S, Hirt J, Homborg A, Köberich S, Kuhn R, Rommel A, Sonnberger M; DEWISS network. Delphi studies in social and health sciences-Recommendations for an interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. *PLoS One*. 2024;19(8):e0304651. DOI: 10.1371/journal.pone.0304651
36. Linstone HA, Turoff M. The Delphi method: Techniques and applications. Addison-Wesley: Reading, Mass; 1975.
37. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm*. 2016;38(3):655-662. DOI: 10.1007/s11096-016-0257-x
38. Shewade HD, Jeyashree K, Kalaiselvi S, Palanivel C, Panigrahi KC. Competency-based tool for evaluation of community-based training in undergraduate medical education in India - a Delphi approach. *Adv Med Educ Pract*. 2017;8:277-286. DOI: 10.2147/AMEP.S123840
39. Niederberger M, Köberich S; members of the DeWiss Network. Coming to consensus: the Delphi technique. *Eur J Cardiovasc Nurs*. 2021;20(7):692-695. DOI: 10.1093/eurjcn/zvab059
40. Spranger J, Niederberger M. Handbuch: Unipark für Delphi-Befragungen. Schwäbisch Gmünd: University of Education; 2023. Zugänglich unter/available from: https://phsg.bsz-w.de/frontdoor/deliver/index/docId/455/file/Handbuch_unipark_Delphi-Befragungen_SoSe23.pdf
41. Aengenheyster S, Cuhls K, Gerhold L, Heiskanen-Schüttler M, Huck J, Muszynska M. Real-Time Delphi in practice — A comparative analysis of existing software-based tools. *Technol Forecast Soc Change*. 2017;118:15-27. DOI: 10.1016/j.techfore.2017.01.023
42. Niederberger M, Renn O. Das Gruppendelphi-Verfahren. Wiesbaden: Springer Fachmedien; 2018. DOI: 10.1007/978-3-658-18755-2
43. Groetker R. Expertenkonsultationen und Stakeholder-Befragungen mit Deliberativem Delphi (Expert Consultations and Stakeholder-Surveys with Deliberative Delphi). SSRN Electronic J. 27.09.2017. DOI: 10.2139/ssrn.3256258
44. Estrella Porter PD, Ayala Mullo JF, Barba Carrera DA, Barros Castro AX, Cabascango Vasquez ES, Del Castillo Arellano JC, Condo Espinel PX, Eid Arellano EG, Estrella Porter JA, Falconi Paez AC, Fierro Valle MC, Gallegos Miranda PJ, Guerra Velastegui AR, Guevaraz Baez DC, Jara Santamaria BD, Lopez Diaz JA, Mejia Viana JC, Moya Quitto GF, Muenala PM, Muenala TC, Nocolalade López Bi, Oquendo Carrera AD, Ordoñez Paz AL, Ortiz Duque EF, Palacios Granda MC, Patoja Borja NS, Puga Martinez SE, Rueda Ordoñez CJ, Soto Gutierrez LP, Tixi Tapia GE, Torres Moscoso MB, Vaca Porras MA, Viteri Suárez MI, Gullemot JR. Medical education from the point of view of medical students: Results from four participatory Delphi panels in Quito, Ecuador. *Med Teach*. 2020;42(9):1051-1057. DOI: 10.1080/0142159X.2020.1792865
45. Copeland C, Fisher J, Teodorczuk A. Development of an international undergraduate curriculum for delirium using a modified delphi process. *Age Ageing*. 2018;47(1):131-137. DOI: 10.1093/ageing/afx133
46. Salihu HM, Dongarwar D, Malmberg ED, Harris TB, Christner JG, Thomson WA. Curriculum Enrichment Across the Medical Education Continuum Using e-Delphi and the Community Priority Index. *South Med J*. 2019;112(11):571-580. DOI: 10.14423/SMJ.0000000000001033
47. Duncan E, O'Cathain A, Rousseau N, Croot L, Sworn K, Turner KM, Yardley L, Hoddinott P. Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. *BMJ Open*. 2020;10(4):e033516. doi: 10.1136/bmjopen-2019-033516
48. Spickermann A, Zimmermann M, von der Gracht HA. Surface- and deep-level diversity in panel selection — Exploring diversity effects on response behaviour in foresight. *Technol Forecast Soc Change*. 2014;85:105-120. DOI: 10.1016/j.techfore.2013.04.009
49. Baker J, Lovell K, Harris N. How expert are the experts? An exploration of the concept of 'expert' within Delphi panel techniques. *Nurse Res*. 2006;14(1):59-70. DOI: 10.7748/nr2006.10.14.1.59.c6010
50. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs*. 2000;32(4):1008-1015.
51. Turoff M. The Policy Delphi. In: Linstone HA, Turoff M, editors. The Delphi method: Techniques and applications. Newark (NJ): New Jersey Institute of Technology; 2002. p.80-96.
52. Viljoen CA, Millar RS, Manning K, Burch VC. Determining electrocardiography training priorities for medical students using a modified Delphi method. *BMC Med Educ*. 2020;20(1):431. DOI: 10.1186/s12909-020-02354-4
53. Garavalia L, Gredler M. Teaching Evaluation through Modeling: Using the Delphi Technique to Assess Problems in Academic Programs. *Am J Eval*. 2004;25(4):375-380. DOI: 10.1016/j.ameval.2004.05.006
54. Lei L, Lin L, Deng J, Dong H, Luo Y. Developing an Organ Donation Curriculum for Medical Undergraduates in China Based on Theory of Planned Behavior: A Delphi Method Study. *Ann Transplant*. 2020;25:e922809. DOI: 10.12659/AOT.922809
55. Shah S, McCann M, Yu C. Developing a National Competency-Based Diabetes Curriculum in Undergraduate Medical Education: A Delphi Study. *Can J Diabetes*. 2020;44(1):30-36.e2. DOI: 10.1016/j.cjcd.2019.04.019
56. Urushibara-Miyachi Y, Kikukawa M, Ikusaka M, Otaki J, Nishigori H. Lists of potential diagnoses that final-year medical students need to consider: a modified Delphi study. *BMC Med Educ*. 2021;21(1):234. DOI: 10.1186/s12909-021-02652-5
57. Barrios M, Guilera G, Nuño L, Gómez-Benito J. Consensus in the delphi method: What makes a decision change? *Technol Forecast Soc Change*. 2021;163:120484. DOI: 10.1016/j.techfore.2020.120484
58. von der Gracht HA. Consensus measurement in Delphi studies. *Technol Forecast Soc Change*. 2012;79(8):1525-1536. DOI: 10.1016/j.techfore.2012.04.013
59. Lange T, Kopkow C, Lutzner J, Gunther KP, Gravius S, Scharf HP, Stöve J, Wagner R, Schmitt J. Comparison of different rating scales for the use in Delphi studies: different scales lead to different consensus and show different test-retest reliability. *BMC Med Res Methodol*. 2020;20(1):28. DOI: 10.1186/s12874-020-0912-8
60. Del Grande C, Kaczorowski J. Rating versus ranking in a Delphi survey: a randomized controlled trial. *Trials*. 2023;24(1):543. DOI: 10.1186/s13063-023-07442-6
61. Johns R. One Size Doesn't Fit All: Selecting Response Scales For Attitude Items. *J Elect Public Opin Parties*. 2005;15(2):237-264. DOI: 10.1080/13689880500178849
62. Belton I, MacDonald A, Wright G, Hamlin I. Improving the practical application of the Delphi method in group-based judgement: a six-step prescription for a well-founded and defensible process. *Health Technol Assess*. 2019;147:72-82. DOI: 10.1016/j.techfore.2019.07.002
63. Artino Jr AR, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach*. 2014;36(6):463-474. DOI: 10.3109/0142159X.2014.889814

64. O'Cathain A, Thomas KJ. "Any other comments?" Open questions on questionnaires - a bane or a bonus to research? *BMC Med Res Methodol.* 2004;4:25. DOI: 10.1186/1471-2288-4-25
65. Mayring P. Einführung in die qualitative Sozialforschung : eine Anleitung zu qualitativem Denken. 7., überarbeitete Auflage. Weinheim, Basel: Beltz; 2023.
66. Braun V, Clarke V. Thematic analysis: a practical guide. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE; 2022.
67. Fröh W. Inhaltsanalyse: Theorie und Praxis. 9. überarbeitete Auflage. Konstanz: UVK Verlagsgesellschaft mbH; 2017. DOI: 10.36198/9783838547350
68. Niederberger M, Schifano J, Deckert S, Hirt J, Homberg A, Köberich S, Kuhn R, Rommel A, Sonnberger M; DEWISS network. Delphi studies in social and health sciences-Recommendations for an interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. *PLoS One.* 2024;19(8):e0304651. DOI: 10.1371/journal.pone.0304651
69. Schmidt RC. Managing Delphi Surveys Using Nonparametric Statistical Techniques. *Decis Sci.* 2007;28(3):763-774. DOI: 10.1111/j.1540-5915.1997.tb01330.x
70. Dajani JS, Sincoff MZ, Talley WK. Stability and agreement criteria for the termination of Delphi Studies. *Technol Forecast Soc Change.* 1979;13:83-90. DOI: 10.1016/0040-1625(79)90007-6
71. Gattrell WT, Logullo P, van Zuuren EJ, Price A, Hughes EL, Blazey P, Winchester CC, Tovey D, Goldman K, Hungin AP, Harison N. ACCORD (ACcurate CONsensus Reporting Document): A reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLoS Med.* 2024;21(1):e1004326. DOI: 10.1371/journal.pmed.1004326
72. Wallace MJ, Zecharia A, Guilding C, Tucker S, McFadzean I. Developing a new undergraduate pharmacology core curriculum: The British Pharmacological Society Delphi Method. *Pharmacol Res Perspect.* 2021;9(4):e00832. DOI: 10.1002/prp2.832

Korrespondenzadresse:

Angelika Homberg
Medizinische Fakultät Mannheim, Universität Heidelberg,
Geschäftsbereich Studium und Lehrentwicklung, Abteilung
Medizinische Ausbildungsforschung, Theodor-Kutzer-Ufer
1-3, 68167 Mannheim, Deutschland
Angelika.homberg@medma.uni-heidelberg.de

Bitte zitieren als

Homberg A. Conducting Delphi surveys in medical education research. *GMS J Med Educ.* 2026;43(1):Doc6.
DOI: 10.3205/zma001800, URN: urn:nbn:de:0183-zma0018008

Artikel online frei zugänglich unter

<https://doi.org/10.3205/zma001800>

Eingereicht: 30.09.2024

Überarbeitet: 29.04.2025

Angenommen: 05.06.2025

Veröffentlicht: 15.01.2026

Copyright

©2026 Homberg. 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/>.