Zeitschrift für Didaktik der Naturwissenschaften (2024)30:4
https://doi.org/10.1007/s40573-023-00169-6

ORIGINAL PAPER

What Challenges Do Researchers Face in the Study of Inclusive Science Education?

A Delphi Study

Felix Pawlak[1] · Jürgen Menthe[2] · Elizabeth Watts[3] · Lisa Stinken-Rösner[4]

Received: 24 January 2023 / Accepted: 23 November 2023
© The Author(s) 2023

Abstract
The study of inclusive science education poses researchers with new challenges and tasks. There is an inherent complexity within this field as it requires understanding of science subject matter in combination with the normative demands of inclusive pedagogy. A Delphi study was conducted to systematically survey research challenges that arise in the research of inclusive science education from the perspective of experts. In the Delphi study, challenges for research on inclusive science education were identified using an exploratory approach in the first round. In the course of the following rounds, a consensus could be reached for eight challenges (>50% agreement, IQR <1, SD <1.5). These included challenges known from the literature, e.g. a lack of suitable research instruments. The results also showed a change or an overcoming of challenges documented in earlier articles. For example, many experts, no longer experience a lack of research studies as a challenge. Alongside the progress made in the research field of inclusive science education, challenges still need to be addressed, such as the discrepancy between existing theoretical approaches and actual teaching practices. Future research must address these issues.

Keywords Inclusion · Science Education · Delphi Study

✉ Felix Pawlak
felix.pawlak@uni-tuebingen.de

1 Didaktik der Chemie, Eberhard Karls Universität Tübingen, Tübingen, Germany

2 Institut für Biologie und Chemie, Abteilung Chemie, Universität Hildesheim, Hildesheim, Germany

3 Sprachenzentrum, Bauhaus-Universität Weimar, Weimar, Germany

4 Fakultät für Physik, Physik und ihre Didaktik, Universität Bielefeld, Bielefeld, Germany

Published online: 28 December 2023

Vor welchen Herausforderungen stehen Fachdidaktiker:innen bei der Forschung zu inklusivem naturwissenschaftlichen Unterricht?

Eine Delphi-Studie

Zusammenfassung

Forschung zu inklusivem naturwissenschaftlichen Unterricht stellt Forscher:innen vor neue Aufgaben und Herausforderungen. Das Forschungsfeld ist unter anderem deshalb komplex und interdisziplinär, da sowohl die Perspektive der jeweiligen Naturwissenschaftsdidaktik als auch der inklusiven Pädagogik berücksichtigt werden muss. Die durchgeführte Delphi-Studie identifiziert systematisch Herausforderungen, die aus der Sicht von Expert:innen die Forschung auf diesem Gebiet begleiten. In der Delphi-Studie wurden in der ersten Runde mit Hilfe eines explorativen Ansatzes Herausforderungen erfasst. Im Verlauf der weiteren Runden der Delphi-Studie konnte für acht Items ein Konsens (>50 % Zustimmung, IQR <1, SD <1,5) erzielt werden. Dazu gehören aus der Theorie bekannte Herausforderungen, wie das Fehlen geeigneter Forschungsinstrumente. Die Ergebnisse zeigen, dass einige in früheren Artikeln beschrieben Herausforderungen mittlerweile (zum Teil) überwunden werden konnten. So stellt für viele Expert:innen der Mangel an Forschungsarbeiten keine Herausforderung mehr dar. Neben den Fortschritten, die im Forschungsfeld des inklusiven naturwissenschaftlichen Unterrichts gemacht wurden, müssen jedoch noch Herausforderungen angegangen werden, wie die Diskrepanz zwischen den theoretischen Ansätzen und der Umsetzung in der Praxis.

Schlüsselwörter Inklusion · Naturwissenschaftsdidaktik · Delphi-Studie

Introduction

Inclusion aims to enable participation in education for all students (Booth and Ainscow 2002). This includes recognizing the diverse potential of all learners while minimizing barriers (Florian and Spratt 2013; Florian 2014; Mastropieri and Scruggs 2014). As the diversity of the students increases, so do the requirements to meet the needs of the learning group. This poses new challenges for the entire school system and particularly for subject-specific learning and educational research.

Science subject matter is already highly complex due to its constructs, theoretical frameworks, and scientific methodology (de Carvalho 2016; Johnstone 1991). For example, findings from experiments need to be explained. Theories and models are used to make predictions (Marniok and Reiners 2017). Many models used in science are highly complex in themselves. For example, atomic models (Adesokan 2015; Pawlak and Groß 2021) or electric circuit models can be very challenging, as known from research on student conceptions (Müller and Schecker 2018; Wilhelm and Hopf 2018).

At the same time, course content needs to be diversified and, at times, simplified to meet the needs of all learners according to the principles of inclusive pedagogy. In conclusion, science education needs to be reconceptualized considering the challenges arising from the wide diversity students (de Carvalho 2016). A large driver of this reconceptualization is the research conducted on inclusive science education, which examines the efficacy of particular methodologies. However, this field of research raises new tasks and challenges, which need to be systematically documented and addressed.

Theoretical Background

Since the *Salamanca statement* (UNESCO 1994), providing education for all has become a recognized human right. Consequently, the focus has shifted from debating the necessity of inclusive education to exploring effective implementation strategies at all levels (Booth and Ainscow 2002). The topic of inclusion has long been neglected in subject-specific teaching and learning, such as science education (Abels and Stinken-Rösner 2022). There have been repeated calls for the implementation of inclusion to be addressed from within subject-specific discourses (Seitz 2006).

Research on inclusive science education initially aimed to identify challenges in implementing inclusion in the science classroom (Markic and Abels 2014). For example, challenges science teachers face when teaching students with special needs (Abels 2015; Adesokan 2015; Freedberg et al. 2019; Grumbine and Alden 2006; Teke and Sozbilir 2019), or challenges of conducting science experiments in inclusive classrooms (Pawlak and Groß 2021).

The substantial lack of empirical studies and publications on inclusive science education that was initially lamented by researchers (Adesokan 2015; Bianchirli and Cavazos 2001; Blumberg and Mester 2017; Spencer and Marschark 2010; Schlüter 2018; Menthe and Hoffmann 2015) has decreased since the *UN Convention on the Rights of Persons with Disabilities* in 2006 (Brauns and Abels 2020). Today, various approaches exist for planning inclusive science lessons (Schlüter et al. 2016; Schlüter and Melle 2017; Spaulding and Flannagan 2012; Roski et al. 2021), some of which explicitly address subject-specific learning in combination with inclusive pedagogy (Abels and Stinken-Rösner 2022; Stinken-Rösner et al. 2020).

When conducting studies on inclusive science education, researchers face a complex, interdisciplinary, and novel research field (Menthe and Hoffmann 2015; Schlüter 2018). For example, the target group and the associated understanding of inclusion must be clarified (Reiners and Adesokan 2017; Schlüter 2018). When using the narrow understanding of inclusion that focuses on students with special educational needs, a well-known challenge is the small sample sizes (Baumann and Melle 2019). The use of methods based on quantitative data is thus limited. Another challenge arises when individual learners are the focus of research studies, as this may overshadow the subject-specificity of science education (Abels and Stinken-Rösner 2022; Pawlak et al. 2023). Köpfer (2021) expands the differentiation between a narrow understanding of inclusion and a broad understanding that considers all relevant dimensions of diversity by further distinguishing between an individual and an organizational perspective. Inclusion thus appears as a point of tension between two perspectives: the consideration of individual characteristics and the consideration of educational institutions and structures. A challenge for educational research is therefore to refrain from reinforcing exclusion by categorization and labeling of students (Katzenbach 2017).

One approach to support inclusive science teaching is the so-called *NinU-framework*. It was developed collaboratively by researchers from the Network for Inclusive Science Education (NinU) and systematically combines two perspectives: inclusive pedagogy and science education. Regarding inclusive pedagogy, the authors propose three steps in the planning (and reflecting) of science education to make science content accessible and relevant for all learners: valuing diversity, recognizing barriers, and enabling participation (Stinken-Rösner et al. 2020). In the *NinU-framework* these three steps are applied to four commonly agreed goals of science teaching (derived and slightly adapted from the four goals of science teaching by Hodson 2014): reasoning about scientific contexts, learning scientific content, doing science, learning about science.

However, the challenges in the study of inclusive science education are mostly mentioned indirectly, such as the limitations of studies. To facilitate future research in inclusive science education, these challenges need to be identified more clearly. There are hardly any documented reports by scholars on how research on inclusive science education differs from the research on science education in general or research on inclusive education in other subjects. One reason for the latter is certainly the specialization of researchers on individual subjects. We must first identify the subject-related challenges in research on inclusive education to work out the commonalities and specifics of the individual subjects subsequently. This paper aims to take a first step in this direction by presenting the results of a Delphi study on the challenges of research on inclusive science education.

Aims and Research Questions

The aim of this study is to support researchers in future projects by identifying specific challenges for the research field of inclusive science education and making them transparent to the community. So far, however, only a few challenges have been mentioned in publications (Sect. 2). This discrepancy between the complexity of the research field, which has to consider the perspectives of science education and inclusive pedagogy in equal measure, and the low number of explicitly named challenges in researching, is striking and represents a research desideratum. Accordingly, the first research question is:

(1) What are current challenges in the study of inclusive science education?

Based on the results, similarities, and differences to challenges in research on science education in general are discussed.

The second research question examines potential ways of fostering research on inclusive science education, exemplified using the so-called *NinU-framework* (Stinken-Rösner et al. 2020).

(2) To what extent does the *NinU-framework* help support research in the field of inclusive science education?

The aim of the research question (2) is to determine whether existing tools like the *NinU-framework* can help to overcome the challenges identified specifically for research on inclusive science education. In order to answer both research questions, a Delphi study was conducted. In this paper, however, the research question (1) results are presented as a matter of priority. The results of research question (2) are presented solely in the online material. Corresponding background information on the *NinU-framework* can be found in Stinken-Rösner et al. (2020).

Design of the Study

Figure 1 gives an overview of the presented Delphi study. Overall, the study consisted of three rounds of surveys over the course of seven months.

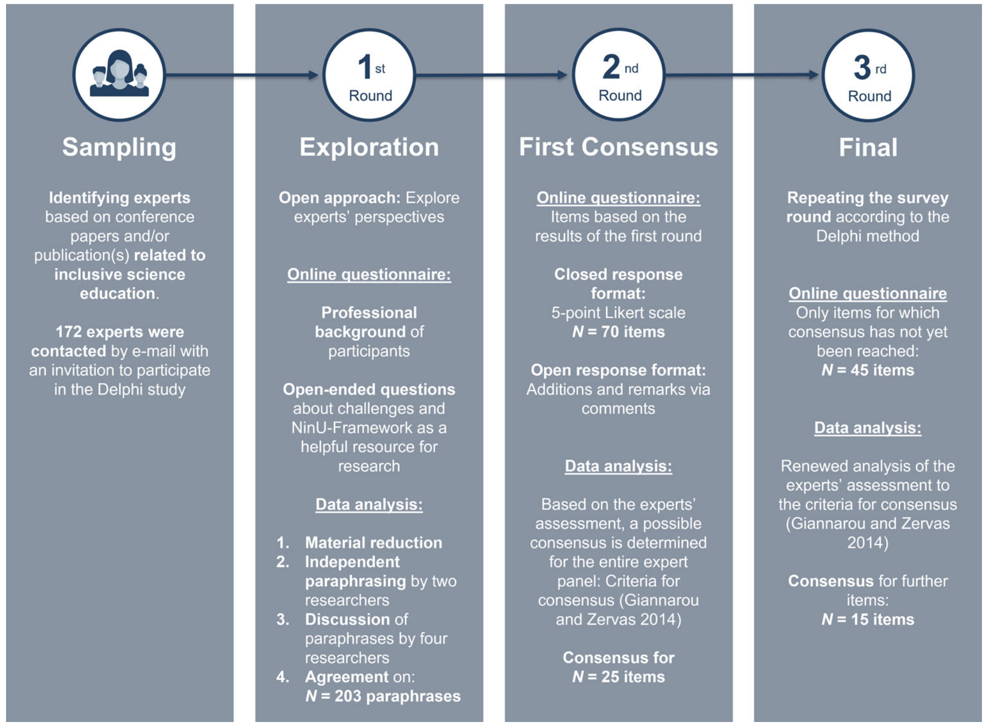

Fig. 1 Project overview

All rounds of the Delphi study were conducted via online questionnaires using LimeSurvey Version 3.28.17 (first round) and EvaSys V8.2 (second and third rounds). The experts participated in all rounds of the study anonymously and with personal codes.

Exploration

The first round of the Delphi study aimed to investigate experts' experiences in this research area. The online questionnaire consisted of two sections: The first part included questions about the participants' professional background (current position, scientific discipline, years of experience in research on inclusive science education, and experience with the *NinU-framework*) to determine their status for the expert panel. The second part addressed the two research questions. Two open-ended questions are used: (1) *What do you find to be particularly challenging in conducting research on inclusive science education?* and (2) *In your opinion, is the NinU-framework a helpful resource for research on inclusive science education?*

Following the data collection, participants' free-text responses were analyzed using qualitative content analysis (Mayring 2015). The core statements were extracted through paraphrasing by two researchers. Where necessary, discursive consensus building was carried out using the original material. By doing so, the scope of the original material could be reduced and clarified (Mayring 2015).

In the next step, the paraphrases were further abstracted, and semantically identical paraphrases were identified and combined (Mayring 2015). The collected material was then compiled into key statements. The resulting sets of paraphrases formed the basis for the items for the following rounds of the Delphi study, respectively. One representative item was formulated for each category. In the last step, the items were put into a logical order along with the original questions to which the experts responded in the first Delphi round.

First Consensus and Uncertain Cases

In the second round of the Delphi study, a questionnaire based on the items derived from the first round was used. Just as in the first round, the questionnaire consisted of two sections: questions regarding the professional background and the same two research questions as in the first round. Experts were presented with items derived from the first round as possible responses and asked to rank their (dis)agreement on a 5-point Likert scale, ranging from 1: *strongly disagree* to 5: *strongly agree*. Additionally, a comment field was included for each item so that experts could explain their ranking. Again, experts were invited via email to participate within three weeks. All experts received a graphic summary of the procedures for expert identification, data collection, analysis process, and the results of the first round together with the invitation.

In analyzing the data, three statistical measurements were performed simultaneously to identify consensus (Giannarou and Zervas 2014):


A challenge/challenges in research on inclusive science education (ISE) is/are…

2.1 …the lack of experts.

Strongly disagree ☐ ☐ ☐ ☐ ☐ Strongly agree

Results of 2nd round: 37,5 % disagreement ↔ 47,5 % agreement | SD = 1,23 | IQR = 2

Comment:

Fig. 2 Example item with results from the previous round

1. The majority (>50%) of experts (dis)agree with the statement. This corresponds to the values 1 or 2 for disagreement and 4 or 5 for agreement (Loughlin and Moore 1979; McKenna 1994; Hackett et al. 2006),
2. the interquartile range of the rankings is below 1.0 (Raskin 1994; Rayens and Hahn 2000), and
3. the standard deviation of the rankings is below 1.5 (Christie and Barela 2005).

The methodological procedure of the third round of the Delphi study was identical to that of the second round. In order to reduce experts' workload, only those items were included for which no consensus had been reached (Scholles 2008). The results from the second round were presented next to each item (see Fig. 2). That allowed experts to reconsider their responses using the feedback from the previous round (Kallia et al. 2021).

The survey concluded after the third round. This decision was based on the reasoning that possible challenges strongly depend on scholars' unique work environments and available resources. Therefore, challenges might be relevant for some but irrelevant for other. In this case, an iterative process would never come to a consensus. A second reason was the drop in the number of participants each round, which is not uncommon. Nevertheless, a Wilcoxon-signed ranks test was conducted, to examine whether stability had been reached which would be indicative of a low probability of achieving consensus in further rounds.

Results

The results are presented in three parts. First, participants and their expert status are described. Then, the results of the three Delphi rounds are summarized.

Participants

The panel of experts was identified based on their contributions and publications in the field of inclusive science education. Specifically, this encompassed (at least one) of the following: a publication in conference proceedings with a focus on inclusive science education (search term: inklusi*) at the major science education conferences in Germany (*DPG, GDCP*, and *FDdB*) between 2017 and 2021, publications in recent books on inclusive science education, participation in *NinU*-meetings (regular meetings of the network for scientific exchange), and/or specific recommendations made by other identified experts.

In total, 196 experts were identified. Due to the limitation of the study to the German-speaking community, seven experts were excluded from the sample. No current contact data could be found for a further 17 experts. The panel was thus reduced to 172 experts.

For each Delphi round, experts were contacted via email with the request to participate. In the course of the study, participation declined from 80 experts in the first round to 29 experts in the third round (Fig. 3).

The majority of experts are specialized in biology, chemistry, or physics education. Experience in research on inclusive science education varied with professional positions. While (trainee) teachers and PhD students stated that they were active in this field between one to five years, some scientific associates and professors reported more than five years of experience. Note that these distributions varied across the Delphi rounds, a detailed description of the sample can be found in the online material.

Round 1

Experts' free-text-answers (N = 121) were reformulated into N = 203 paraphrases, of which 113 paraphrases were in response to question (1) and 90 paraphrases to question (2). The higher number of paraphrases is due to the fact that some free-text answers were divided into several paraphrases, each addressing a different point. Paraphrases

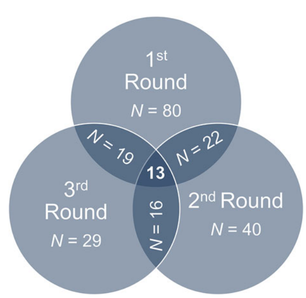

Fig. 3 Participation of experts

Table 1 Example of the item-generation process

Original	Paraphrase	Item
"A challenge/challenges in research on inclusive science education is/are ..."		
"Approval procedures (obtaining permission from parents, ethics committee, etc.) are particularly challenging for studies in inclusive classes" (U-37-T)	... the approval procedures for studies in inclusive classes	**... the approval procedures**
"The permission to interview teachers in our state is very difficult to achieve [...]" (T-82-N)	... the approval for interviewing teachers	
"The biggest challenge is to get approval for the methods to collect data. The combination of video and audio data seems to be the most valuable. This is rarely approved at the moment or requires too long a lead time" (N-30-N)	... the approval of data collection methods (e.g., video & audio data)	

Table 2 Overview of the individual rounds of the Delphi survey

	1st Round	2nd Round	3rd Round
Professional Background	✓	✓	✓
Q (1): *"What do you find to be particularly challenging in conducting research (planning, implementing, and/or evaluating) on inclusive science education?"*	Open-ended, 61 free-text-answers, 133 paraphrases grouped into 26 categories	26 items, consensus for 2	24 items, consensus for 6
Feedback on the results with comment function	X	✓	✓

were grouped into categories, each represented by one prototypical item. An example of the item-generation process is shown in Table 1.

Overall, 70 items were derived from the original material. The final set of items addressed three areas of interest:

1. challenges in research on inclusive science education ($N_{\text{Challenges}} = 26$),
2. the suitability of the *NinU-framework* for research on inclusive science education ($N_{\text{NinU-framework}} = 25$),
3. the suitability of the *NinU-framework* for lesson planning as part of research studies ($N_{\text{Lesson Planning}} = 19$).

In order to ensure objectivity and validity in the item generation process, subsequent quality control was carried out for all experts who participated in both rounds (N= 22). For this purpose, an external researcher, who was not involved in paraphrasing or formulating the items, was presented with the experts' open-ended responses (first round), and asked to rate the items (second round) accordingly. In 95.3% of the cases, the experts themselves (in the second round) and the external researcher (based on the experts' open-ended responses) rated the items analogously with an allowable tolerance of ±1 for ranking.

Rounds 2 and 3

In the second round, consensus was reached for a total of 25 of the 70 items, two relating to question (1), and 23 to question (2). The remaining 45 items were carried over to the third round, where consensus was reached for 15 further items, six relating to question (1), and nine to question (2). An overview of the three rounds is shown in Table 2.

Overall, a consensus was reached for 40 out of 70 items. Additionally, participants used the comment function 107 times in the second round and 68 times in the third round. After the third round, the Delphi study was concluded although consensus could not be reached for all items. However, the experts' responses were stable for most items, as shown by a Wilcoxon-signed rank test (Seagle and Iverson 2002). Only the assessment of items 1.2 ($Z = -2.64$, $p = 0.008$), 1.12 ($Z = -2.08$, $p = 0.047$), and 1.13 ($Z = -2.13$, $p = 0.039$) was significantly different between the second and third rounds. This can also be seen as a criterion for concluding the study (Dajani et al. 1979). The effort of additional rounds would not have been justified, as it was highly unlikely that consensus would have been reached on further items.

Items and results for the research question (2) are presented solely in the online material since this is only a subresult that was not explicitly addressed in the research questions. Table 3 presents the quantitative data. This is followed by an explanation of the (lack of) consensus along the qualitative data (experts' comments).

Consensus on the Challenges in Research on Inclusive Science Education

From experts' point of view, eight aspects are particularly challenging when conducting research on inclusive science education. In the following, explanations for each challenge

Table 3 Challenges of conducting research on inclusive science education (ISE): the highlighted values meet the criteria for consensus

1st Round		2nd Round					3rd Round					Consensus
Item: A challenge/challenges in research on inclusive science education (ISE) is/are		N	SD	IQR	% Disagree	% Agree	N	SD	IQR	% Disagree	% Agree	
1.1	... the lack of experts.	40	**1.23**	2	37.5	47.5	29	**1.22**	2	31	**58.6**	No
1.2	... the small number of existing research studies.	40	**1.17**	2	35	42.5	29	**1.34**	2	44.8	44.8	No
1.3	... the inconsistent definition of inclusion.	40	**1.33**	2.25	35	**52.5**	29	**1.20**	2	20.6	**58.6**	No
1.4	... the different definitions of terms between the scientific disciplines.	39	**1.06**	2	33.3	30.8	29	**1.18**	2	37.9	27.6	No
1.5	... the often negative attitude towards inclusion.	40	**1.28**	3	47.5	30	29	**1.23**	2	48.3	27.6	No
1.6	... the low relevance of the topic of inclusion in science education research.	40	**1.35**	1.25	22.5	47.5	29	**1.23**	2	20.7	**65.5**	No
1.7	... the complex interplay of the perspectives of science education and inclusive pedagogy in research.	40	**1.20**	2	17.5	**67.5**	29	**1.04**	2	6.8	**72.4**	No
1.8	... to consider the great diversity of learners in research.	40	**1.14**	**1**	17.5	**77.5**	–	–	–	–	–	**2nd round**
1.9	... not to lose sight of students with special educational needs in the broad understanding of inclusion.	39	**1.20**	2	30.8	41	28	**1.20**	2	39.3	35.7	No
1.10	... not to fall back into a narrow understanding of inclusion when a broad understanding is desired.	39	**1.25**	1.5	25.6	**56.4**	29	**1.09**	**1**	17.2	**72.4**	**3rd round**
1.11	... to capture the complexity of ISE in research.	39	**1.30**	2	18	**64.1**	29	**1.22**	**1**	17.2	**75.8**	**3rd round**
1.12	... to transfer research results on heterogeneity and/or differentiation to inclusion.	39	**1.10**	2	41	30.8	29	**1.09**	2	41.4	27.5	No
1.13	... working in multi-professional teams.	39	**1.24**	2	43.5	33.3	29	**1.13**	2	**65.5**	10.3	No
1.14	... the acquisition of participants.	40	**1.22**	1.5	25	**57.5**	29	**1.24**	2	24.1	**51.7**	No
1.15	... the approval procedures.	40	**1.17**	**1**	22.5	50	29	**1.30**	2	20.6	**55.2**	No
1.16	... small samples.	40	**1.22**	1.25	25	**57.5**	29	**1.18**	2	20.6	**58.6**	No

Table 3 (Continued)

1st Round		2nd Round					3rd Round					Consensus
Item: A challenge/challenges in research on inclusive science education (ISE) is/are		N	SD	IQR	% Disagree	% Agree	N	SD	IQR	% Disagree	% Agree	
1.17	... the small number of classes with learners with special educational needs.	39	**1.20**	**1**	46.1	23.1	29	**1.40**	2	41.3	27.5	No
1.18	... to focus on appropriate variables in research despite the multiple perspectives on inclusion (e.g. diversity dimensions) and science education.	40	**1.22**	2	15	**70**	29	**1.26**	2	13.7	**69**	No
1.19	... to develop the appropriate instruments.	40	**1.25**	2	17.5	**65**	28	**1.20**	**1**	17.9	**78.5**	**3rd round**
1.20	... the validity of test instruments in the light of different means of access and abilities of the students.	40	**1.04**	2	10	**67.5**	29	**0.99**	**1**	6.9	**75.9**	**3rd round**
1.21	... the division of intervention and control groups.	40	**1.31**	1.25	25	47.5	29	1.54	3	31	**55.1**	No
1.22	... the discrepancy between theoretical approaches and teaching practice.	39	**1.21**	**1**	20.5	**66.7**	–	–	–	–	–	**2nd round**
1.23	... that many different data are needed to capture inclusion in science education.	40	**1.38**	3	30	**62.5**	29	**1.22**	2	17.2	**68.9**	No
1.24	... the dilemma between the formation of categories for (e.g. quantitative) research and the effort to avoid categories (e.g. stereotypes) in the course of inclusion.	39	**1.28**	2	15.4	**69.2**	29	**1.04**	**1**	10.3	**75.8**	**3rd round**
1.25	... to generalize for research purposes and at the same time take individuality into account.	39	**1.16**	2	12.8	**71.8**	29	**1.03**	2	10.3	**72.4**	No
1.26	... the contradiction of an ISE in a selective school system.	40	**1.41**	3	30	**57.5**	28	**0.98**	**1**	7.2	**89.3**	**3rd round**

are derived from the experts' comments from the second and third rounds of the Delphi study.

Challenge 1: To Consider the Great Diversity of Learners The experts distinguish two aspects: On the one hand, they describe it as challenging to capture the diversity of learners (N-39-N; S-18-M; A-43-F)[1], enumerating the various special needs and special talents (A-95-N), but especially the very different individual learning prerequisites (T-21-L; R-30-F; N-40-G). On the other hand, special features of the science education reference disciplines such as differentiation in experimentation (A-81-M) or the fit to content-related or thematic focal points (curricular innovation research (N-44-Z)) are also described as challenges.

Challenge 2: Not to Fall Back into a Narrow Understanding of Inclusion when a Broad Understanding Inclusion is Desired Experts consider it challenging to encompass the wide array of diversity facets in research (N-35-G). Despite shared definitions, implementation in research often reverts to a narrow understanding of inclusion due to the easier measurement of diversity in terms of special education needs (B-39-H; N-35-G). If a broad understanding of inclusion is to be pursued consistently in research, defining the target group is challenging (L-76-E). Scholars need to go beyond seeing heterogeneous learning groups that can be classified according to certain characteristics (N-25-M).

Challenge 3: To Capture the Complexity of ISE in Research The experts explain this challenge by referring to the multidimensionality and complexity of the research object (B-39-H; U-23-R), along with the subsequent issue of presenting research related to this research object concisely and comprehensively in grant applications, while also demonstrating methods that capture this complexity (I-37-T) and further developing them (R-17-L).

Challenge 4: To Develop the Appropriate Instruments Experts emphasize the necessity of developing appropriate research instruments (E-29-T; M-45-E; T-82-N). Designing test instruments accessible to all learners poses a particularly daunting challenge (N-35-G), especially regarding language or comprehension abilities (R-17-L). The goal is to ensure that the instruments accommodate all students' diverse needs and abilities, including those with special needs. Despite these difficulties, there is an optimistic belief that this challenge can be successfully addressed (R-17-L).

Challenge 5: The Validity of Test Instruments in the Light of Different Means of Access and Abilities of the Students Experts describe the difficulty of taking the increased diversity into account in methodological terms. Suitable survey instruments have yet to be found (R-47-D) or developed that meet the same validity standards of established approaches (D-30-B).

Challenge 6: The Discrepancy Between Theoretical Approaches and Teaching Practice Experts recognize a gap between theoretical approaches, often developed in academic settings, and their practical feasibility in teaching practice (M-45-E). While theoretical frameworks and concepts for inclusive science education do exist, there is still a lack of concrete and actionable ideas for implementing these approaches effectively in the classroom (A-85-N). Furthermore, many existing concepts and theories are often deemed unrealistic or challenging to apply in real educational settings (A-58-N; E-89-T). Therefore, the high theoretical goals rarely find their way into teaching practice, making research on inclusive science education even more difficult.

Challenge 7: The Dilemma Between the Formation of Categories for (e.g. Quantitative) Research and the Effort to Avoid Categories (e.g. Stereotypes) in the Course of Inclusion The comments reflect that methodological considerations and normative consequences of certain methodological decisions are interconnected. The use of categories increases the likelihood that research will focus on individual characteristics rather than instructional design (E-47-N), which is important and possible to prevent (I-00-N). The necessity of grouping could potentially contribute to discrimination in the worst case (D-30-B). Similarly, there is an expressed concern that designing materials with specific students in mind may reinforce group differences rather than promote inclusive learning (E-47-N). Quantitative statements could result in discrimination when the goal of comparability (N-35-D) is linked to categorizing students (E-47-Z). The most comprehensive comment points out that while most studies investigate inclusive science education, the research itself is not inclusive (I-69-B). T-82-N highlights that the challenge mentioned here is one of the few dilemmas that are generally true for all research on inclusive science teaching.

Challenge 8: The Contradiction of Inclusive Science Education in a Selective School System The experts describe a particularly drastic situation concerning this challenge. Research on inclusive science education takes place within a school and social system that experts describe as hostile to inclusion (E-89-T). For example, school students are categorized into different types of school (selectivity). Current research focuses on the micro level of classroom practice where inclusion can be partially implemented (E-89-T). It should be noted that this strongly depends on the

[1] The reference indicates the personal code of the expert.

respective school and that positive examples also exist (R-17-L).

Challenges for Some Experts

In addition to the challenges where there was consensus, other challenges did exhibit a clear trend but did not meet the rigid criteria set for consensus (see Sect. 4.2.). For example, items 1.6, 1.7, 1.18, 1.23, and 1.25. all achieved (dis)agreement above 65%, for which the interquartile range was too large for consensus.

65.5% of the experts in the third round agreed with item 1.6 that a challenge lies in the low relevance of the topic of inclusion in science education research. However, a large enough interquartile range indicated that there was also disagreement. The result for item 1.2 is consistent with this finding, with two equally large groups opposing each other: 44.8% of the experts regard the fact that there is not yet enough research in this area as a challenge. Simultaneously, 44.8% of the experts disagreed with this statement. One commenter indicated that much work already exists but there is a lack of a review of that work (R-17-L). Similarly, the expert ratings split into two groups for item 1.9, not to lose sight of students with special educational needs in the broad understanding of inclusion. In addition, special needs educators trained to work with students with certain diagnosed special needs may have their own perspectives here.

The findings on item 1.7, the complex interplay between the perspectives of science education and inclusive pedagogy in research, and item 1.13, working in multi-professional teams, are interesting because they seem contradictory. A clear majority (72.4%) of the experts agree that considering the interplay of two perspectives is a challenge. However, this challenge does not seem to be linked to working in multi-professional teams (65.5% disagreement). In one comment, L-76-E calls for going even one step further and forming not only multi-professional but also inclusive teams.

Additional items, for which a clear majority of the experts voiced agreement, revolve around research methodological challenges: A challenge is seen in attempting to focus on appropriate variables in research despite the multiple perspectives on inclusion (e.g. diversity dimensions) and science education (item 1.18, 69%). Other challenges are that many different sets of data are needed to capture inclusion in science education (item 1.23, 68.9%). It is also difficult to reach generalizations in research while simultaneously taking individuality into account (item 1.25, 72.4%). All three challenges are related to the need to consider multiple perspectives, which in turn need to be further differentiated, making it difficult to draw generalized conclusions. In the comments, “for me not a specific problem of inclusive science education, but difficult in general when researching education” (I-00-N) and “applies to other research as well” (R-17-L), experts raise the fundamental question of whether all challenges are actually unique for research on inclusive science education.

Discussion

In the following, the results and limitations of the study are discussed.

Challenges in Research on Inclusive Science Education

The presented study contributes to the systematic collection and documentation of challenges specific to research on inclusive science education by synthesizing and explaining scholars’ experiences. Overall, the experts agreed on eight challenges, some of which are already published, such as the lack of suitable research instruments (items 1.19 & 1.20; Baumann and Melle 2019). Developing appropriate test instruments that deliver valid results based on a diverse student body with different modes of access and capabilities is still a desideratum. Also, the tension between embracing student diversity (item 1.8) and the desire to capture the complexity of science education in research (item 1.11) remains unresolved (Abels and Stinken-Rösner 2022; Pawlak et al. 2023). Researchers are challenged to navigate between the formation of categories for research and the effort to avoid categories in the course of inclusion (item 1.24; Abels and Stinken-Rösner 2022; Katzenbach 2017). Even though categories may be indispensable for specifically planning and implementing subject lessons and for determining their success (Abels and Stinken-Rösner 2022), inclusive principles should be consistently considered in practice and research (Nind and Vinha 2014). Being transparent about the underlying understanding of inclusion in research can help ensure that researchers do not shift from one definition to another, as was described as a challenge by experts (item 1.10).

In addition to the known challenges, experts highlight new (or previously undocumented) ones as well. For example, the discrepancy between existing theoretical approaches for inclusive science education (Abels and Stinken-Rösner 2022; Schlüter et al. 2016; Schlüter and Melle 2017; Spaulding and Flannagan 2012; Stinken-Rösner et al. 2020; Roski et al. 2021) and teaching practice (item 1.22), as well as the societal contradiction of inclusive science education in mainstream school systems (item 1.26). It seems that the previously lamented lack of research on inclusive science education (Adesokan 2015; Bianchirli and Cavazos 2001; Blumberg and Mester 2017; Spencer and Marschark 2010; Schlüter 2018; Menthe and Hoffmann 2015) has been addressed by the community, but that the developed approaches are still difficult to implement in practice.

However, it also appears that some of the challenges known from the literature have been successfully overcome. For example, the small number of existing research studies (item 1.2; Blumberg and Mester 2017), the inconsistent definition of inclusion (item 1.3; Reiners and Adesokan 2017; Schlüter 2018), and small sample sizes (items 1.16 & 1.17; Baumann and Melle 2019) are no longer perceived as challenges by most experts. The research efforts of recent years and the increasing number of publications (Brauns and Abels 2020, 2021) may have contributed to the perception that these challenges have been mastered.

Interestingly, most of the identified challenges are not specific to research on inclusive science education. In many challenges, the problems tend to be attributed to individuals and their particularities rather than methodological or subject-specific aspects. This aligns with the often-prevailing deficit-oriented attitude toward inclusion criticized by Köpfer (2021). Locating the challenge in the research methodology, in subject-specifics or in institutional limitations rather than in the students could help to take on a more potential-oriented perspective on diversity in terms of a broad understanding of inclusion also in research.

Limitations

One study limitation is the variation of and decreasing number of participants in each round. All of the identified experts were contacted in each round of the Delphi study since there was no matching between participation in one of the previous rounds and the experts' contact information. Therefore, experts could participate in either one, two, or all three rounds. The different participants could have influenced consensus building as they assessed the items differently. In particular, the number of experts in the last round was relatively small. Furthermore, only thirteen experts participated in all three rounds. If only this subsample had been considered, consensus would have been reached for ten additional challenges—most of which were discussed in the Sect. 5.3.2 as items with a clear trend. However, the rigid criteria for consensus prevented the possible dominance of a subgroup. Even when the threshold for the measure of (dis)agreement is increased from 50 to 66%, the presented results remain unchanged.

Another limitation is that no consensus was reached on 18 items regarding the first research question. Some of these challenges seem to be controversial. This may be due to different experiences in research. For example, different research locations offer different conditions. However, one important result is that there is not a clear opinion about some challenges in research on inclusive science education.

One methodological limitation is the implementation of the first Delphi round. A first exploration of challenges was conducted via an online questionnaire containing two open questions. Although this method is simple to carry out, writing detailed text is an obstacle, even for experts. It is also not possible, as in group interviews, for example, to go into more detail on statements made by participants and thus obtain more specific answers. However, such an approach would not be compatible with the Delphi study since the experts would have to (partially) give up their anonymity. Thus, individuals could influence the opinion of others, e.g., due to their position. However, this limitation seems to be only partially relevant since several paraphrases could be extracted from single answers (Sect. 4.1.).

Conclusion and Outlook

The Delphi study aimed to identify challenges for research in the field of inclusive science education. This goal was achieved, as experts consensually agreed on eight challenges. This study illustrates some of the specific challenges researchers face in the field, such as capturing the complexity of inclusive science education in research. Other major challenges identified include the need to consider the full range of diversity among learners and the discrepancy between theoretical approaches and teaching practice. In addition, other challenges could be identified that at least some experts consider essential. For example, several experts perceive a challenge in the low interest in the topic of inclusion in science education research.

The intention of identifying challenges is to highlight difficulties that researchers in the field may face, which need to be addressed within the research community. In this way, solutions may be found that contribute to further research and thus to the goal of inclusive science education.

Supplementary Information The online version of this article (https://doi.org/10.1007/s40573-023-00169-6) contains supplementary material, which is available to authorized users.

Acknowledgements We would like to thank all the experts for participating in this study. We would also like to thank *Sarah Hundertmark* for the preliminary work and all *NinU* members for their support and constructive feedback.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abels, S. (2015). Scaffolding inquiry-based science and chemistry education in inclusive classrooms. In N.L. Yates (Ed.), *New developments in science education research* (pp. 77–95). New York: Nova Science Publishers.

Abels, S., & Stinken-Rösner, L. (2022). Inklusion als Phänomen in Chemie- und Physikdidaktik – Gemeinsamkeiten und Unterschiede. In M. Braksiek, K. Golus, B. Gröben, M. Heinrich, P. Schildhauer, & L. Streblow (Eds.), *Schulische Inklusion als Phänomen – Phänomene schulischer Inklusion* (pp. 273–292). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-34178-7_12

Adesokan, A. (2015). *Zur Förderung einer naturwissenschaftlichen Grundbildung bei Schülerinnen und Schülern mit Hörbeeinträchtigung – Eine qualitative Studie als Beitrag zur Entwicklung eines inklusiven Chemieunterrichts*. Universität Köln.

Menthe, J., & Hoffmann, T. (2015). Inklusiver Chemieunterricht: Chance und Herausforderung. In O. Musenberg & J. Riegert (Eds.), *Inklusiver Fachunterricht in der Sekundarstufe* (pp. 131–141). Stuttgart: Kohlhammer.

Baumann, T., & Melle, I. (2019). Evaluation of a digital UDL-based learning environment in inclusive chemistry education. *Chemistry Teacher International*, *1*(2), 20180026. https://doi.org/10.1515/cti-2018-0026.

Bianchirli, J.A., & Cavazos, L.M. (2001). Promoting inclusive science education through professional development: challenges faced in transforming content and pedagogy. *Counterpoints*, *150*, 259–293.

Blumberg, E., & Mester, T. (2017). Potenzielle Gelingensbedingungen für inklusives Lernen im naturwissenschaftlich-technischen Sachunterricht in der Grundschule – Auf dem Weg zu empirischen Evidenzen. In F. Hellmich & E. Blumberg (Eds.), *Inklusiver Unterricht in der Grundschule* (pp. 294–312). Stuttgart: Kohlhammer.

Booth, T., & Ainscow, M. (2002). *Index for inclusion: developing learning and participation in schools*. Bristol: CSIE.

Brauns, S., & Abels, S. (2020). *The framework for inclusive science education*. Inclusive Science Education, Working Paper 1/2020. (pp. 1–126).

Brauns, S., & Abels, S. (2021). Die Anwendung naturwissenschaftlicher Untersuchungsmethoden inklusiv gestalten – Naturwissenschaftsdidaktische Theorie und Empirie erweitern mit dem Kategoriensystem inklusiver naturwissenschaftlicher Unterricht (KinU). *Zeitschrift für Didaktik der Naturwissenschaften* 27(1), 231–249. https://doi.org/10.1007/s40573-021-00135-0.

de Carvalho, R. (2016). Science initial teacher education and superdiversity: educating science teachers for a multi-religious and globalised science classroom. *Cultural Studies of Science Education*, *11*(2), 253–272.

Christie, C.A., & Barela, E. (2005). The Delphi technique as a method for increasing inclusion in the evaluation process. *The Canadian Journal of Program Evaluation*, *20*(1), 105–122.

Dajani, J.S., Sincoff, M.Z., & Talley, W.K. (1979). Stability and agreement criteria for the termination of Delphi studies. *Technological Forecasting and Social Change*, *13*(1), 83–90.

Florian, L. (2014). What counts as evidence of inclusive education? *European Journal of Special Needs Education*, *29*(3), 286–294. https://doi.org/10.1080/08856257.2014.933551.

Florian, L., & Spratt, J. (2013). Enacting inclusion: a framework for interrogating inclusive practice. *European Journal of Special Needs Education*, *28*(2), 119–135. https://doi.org/10.1080/08856257.2013.778111.

Freedberg, S., Bondie, R., Zusho, A., & Allison, C. (2019). Challenging students with high abilities in inclusive math and science classrooms. *High Ability Studies*, *30*, 237–254. https://doi.org/10.1080/13598139.2019.1568185.

Giannarou, L., & Zervas, E. (2014). Using Delphi technique to build consensus in practice. *International Journal of Business Science and Applied Management*, *9*(2), 65–82.

Grumbine, R., & Alden, P.B. (2006). Teaching science to students with learning disabilities. *Science Teacher*, *73*, 26–31.

Hackett, S., Masson, H., & Phillips, S. (2006). Exploring consensus in practice with youth who are sexually abusive: Findings from a Delphi study of practitioner views in the United Kingdom and the Republic of Ireland. *Child Maltreatment*, *11*(2), 146–156.

Hodson, D. (2014). Learning science, learning about science, doing science: different goals demand different learning methods. *International Journal of Science Education*, *36*(15), 2534–2553. https://doi.org/10.1080/09500693.2014.899722.

Johnstone, A.H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, *7*(2), 75–83. https://doi.org/10.1111/j.1365-2729.1991.tb00230.x.

Kallia, M., van Borkulo, S.P., Drijvers, P., Barendsen, E., & Tolboom, J. (2021). Characterising computational thinking in mathematics education: A literature-informed Delphi study. *Research in Mathematics Education*, *23*(2), 159–187. https://doi.org/10.1080/14794802.2020.1852104.

Katzenbach, D. (2017). Inklusion und Heterogenität. In T. Bohl, J. Budde & M. Rieger-Ladich (Eds.), *Umgang mit Heterogenität in Schule und Unterricht* (pp. 124–140). Bad Heilbrunn: Klinkhardt.

Köpfer, A. (2021). Rekonstruktion und Inklusion – Perspektiven und Spannungsfelder rekonstruktiver Inklusionsforschung in der Erziehungswissenschaft. *Qualifizierung für Inklusion*. https://doi.org/10.21248/qfi.64.

Loughlin, K.G., & Moore, L.F. (1979). Using Delphi to achieve congruent objectives and activities in a paediatrics department. *Journal of Medical Education*, *54*(2), 101–106.

Markic, S., & Abels, S. (2014). Heterogeneity and diversity: a growing challenge or enrichment for science education in German schools? *EURASIA Journal of Mathematics, Science & Technology Education*, *10*(4), 271–283.

Marniok, K., & Reiners, C.S. (2017). Representations of nature of science in German school chemistry textbooks. In C.V. McDonald & F. Abd-El-Khalick (Eds.), *Representations of nature of science in school science textbooks: a global perspective* (pp. 201–2140). New York: Routledge.

Mastropieri, M.A., & Scruggs, T.E. (2014). *The inclusive classroom: strategies for effective differentiated instruction*. Boston: Pearson.

Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12th edn.). Weinheim: Beltz.

McKenna, H.P. (1994). The Delphi technique: a worthwhile research approach for nursing? *Journal of Advanced Nursing*, *19*(6), 1221–1225.

Müller, R., & Schecker, H. (2018). Schülervorstellungen zur Quanten- und Atomphysik. In H. Schecker, T. Wilhelm, M. Hopf & R. Duit (Eds.), *Schülervorstellungen und Physikunterricht* (pp. 209–224). Berlin, Heidelberg: Springer Spektrum.

Nind, M., & Vinha, H. (2014). Doing research inclusively: bridges to multiple possibilities in inclusive research. *British Journal of*

Learning Disabilities, *42*(2), 102–109. https://doi.org/10.1111/bld.12013.

Pawlak, F., & Groß, K. (2021). Einsatz von Schülerexperimenten im inklusiven Chemieunterricht – Chancen und Herausforderungen aus Sicht der Chemielehrenden. *CHEMKON* 28(3), 96–102. https://doi.org/10.1002/ckon.201900017

Pawlak, F., Menthe, J., Watts, E., & Stinken-Rösner, L. (2023). Herausforderungen in der Beforschung von inklusivem Nawi-Unterricht. In H. van der Vorst (Eds.), *Lernen, Lehren und Forschen in einer digital geprägten Welt* (pp. 398–401), GDCP-Tagungsband.

Raskin, M.S. (1994). The Delphi study in field instruction revisited: expert consensus on issues and research priorities. *Journal of Social Work Education*, *30*(1), 75–89.

Rayens, M.K., & Hahn, E.J. (2000). Building consensus using the policy Delphi method. *Policy, Politics & Nursing Practice*, *1*(4), 308–315.

Reiners, Ch S., & Adesokan, A. (2017). Inklusion im Chemieunterricht. In Ch S. Reiners (Ed.), *Chemie vermitteln: Fachdidaktische Grundlagen und Implikationen* (pp. 167–177). Springer.

Roski, M., Walkowiak, M., & Nehring, A. (2021). Universal design for learning: the more, the better? *Education Sciences*. https://doi.org/10.3390/educsci11040164.

Schlüter, A.-K. (2018). Professionalisierung angehender Chemielehrkräfte für einen gemeinsamen Unterricht. *Logos*.

Schlüter, A.-K., & Melle, I. (2017). Luft ist komprimierbar. Beispiele für die Umsetzung des Universal Design for Learning. *Naturwissenschaften im Unterricht Chemie*, *162*, 36–39.

Schlüter, A.-K., Melle, I., & Wember, F.B. (2016). Unterrichtsgestaltung in Klassen des Gemeinsamen Lernens. Universal Design for Learning. *Sonderpädagogische Förderung Heute*, *3*(61), 270–285.

Scholles, F. (2008). Delphi. In D. Fürst & F. Scholles (Eds.), *Handbuch Theorien und Methoden der Raum- und Umweltplanung* (pp. 375–380). Dortmund.

Seagle, E., & Iverson, M. (2002). Characteristics of the turfgrass industry in 2020: A Delphi study with implications for agricultural education programs. *Journal of Southern Agricultural Research*, *52*(1), 1–13.

Seitz, S. (2006). Inklusive Didaktik: Die Frage nach dem ‚Kern der Sache'. *Zeitschrift für Inklusion* 1. http://www.inklusion-online.net/index.php/inklusion-online/article/view/184/184. Accessed 25 May 2023.

Spaulding, L.S., & Flannagan, J.S. (2012). Dis2ect: a framework for effective inclusive science instruction. *Teaching Exceptional Children*, *44*(6), 6–14. https://doi.org/10.1177/004005991204400601.

Spencer, P.E., & Marschark, M. (2010). *Evidence-based practice in educating deaf and hard of hearing students*. New York: Oxford University Press.

Stinken-Rösner, L., Rott, L., Hoffmann, T., Menthe, J., Abels, S., Baumann, T., Hundertmark, S., & Nehring, A. (2020). Thinking inclusive science education from two perspectives: inclusive pedagogy and science education. *RISTAL*, *3*, 30–45.

Teke, D., & Sozbilir, M. (2019). Teaching energy in living systems to a blind student in an inclusive classroom environment. *Chemistry Education Research and Practice*, *20*(4), 890–901. https://doi.org/10.1039/c9rp00002j.

UNESCO (1994). *Final report: world conference on special needs education: access and quality*. Paris: UNESCO.

Wilhelm, T., & Hopf, M. (2018). Schülervorstellungen zum elektrischen Stromkreis. In H. Schecker, T. Wilhelm, M. Hopf & R. Duit (Eds.), *Schülervorstellungen und Physikunterricht* (pp. 115–138). Berlin, Heidelberg: Springer Spektrum.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.