Journal of Science Teacher Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uste20

Learning Goals of Professional Development Programs at Science Research Institutions: A Delphi Study with Different Stakeholder Groups

Anja Kranjc Horvat, Jeff Wiener, Sascha Schmeling & Andreas Borowski

To cite this article: Anja Kranjc Horvat, Jeff Wiener, Sascha Schmeling & Andreas Borowski (2021): Learning Goals of Professional Development Programs at Science Research Institutions: A Delphi Study with Different Stakeholder Groups, Journal of Science Teacher Education, DOI: 10.1080/1046560X.2021.1905330

To link to this article: https://doi.org/10.1080/1046560X.2021.1905330

Published online: 28 May 2021.

JOURNAL OF SCIENCE TEACHER EDUCATION
https://doi.org/10.1080/1046560X.2021.1905330

Learning Goals of Professional Development Programs at Science Research Institutions: A Delphi Study with Different Stakeholder Groups

Anja Kranjc Horvat[a,b], Jeff Wiener[b], Sascha Schmeling[b], and Andreas Borowski[a]

[a]Institute of Physics and Astronomy, University of Potsdam, Potsdam, Germany; [b]Teacher and Student Programmes Section, CERN, European Organization for Nuclear Research, Geneva, Switzerland

ABSTRACT
Effective professional development programs (PDPs) rely on well-defined goals. However, recent studies on PDPs have not explored the goals from a multi-stakeholder perspective. This study identifies the most important learning goals of PDPs at science research institutions as perceived by four groups of stakeholders, namely teachers, education researchers, government representatives, and research scientists. Altogether, over 100 stakeholders from 42 countries involved in PDPs at science research institutions in Europe and North America participated in a three-round Delphi study. In the first round, the stakeholders provided their opinions on what they thought the learning goals of PDPs should be through an open-ended questionnaire. In the second and third rounds, the stakeholders assessed the importance of the learning goals that emerged from the first round by rating and ranking them, respectively. The outcome of the study is a hierarchical list of the ten most important learning goals of PDPs at particle physics laboratories. The stakeholders identified enhancing teachers' knowledge of scientific concepts and models and enhancing their knowledge of the curricula as the most important learning goals. Furthermore, the results show strong agreement between all the stakeholder groups regarding the defined learning goals. Indeed, all groups ranked the learning goals by their perceived importance almost identically. These outcomes could help policymakers establish more specific policies for PDPs. Additionally, they provide PDP practitioners at science research institutions with a solid base for future research and planning endeavors.

KEYWORDS
Teacher professional development; Delphi study; multi-stakeholder analysis; pedagogical content knowledge

Introduction

Continuous advances in science and pedagogy are entering the classrooms through curricular reforms (Borko, 2004; Kurnaz & Çepni, 2012; Wallace & Priestley, 2017). There, teachers are often met with demands that go beyond their initial training (OECD, 2019); thus, successful curricular reforms rely considerably on teachers' professional development (Borko, 2004; Corcoran, 1995; Garet et al., 2001; OECD, 2019). Effective professional development is an intentional process with a clear vision of purpose and well-defined goals for teacher learning, student learning, and organization (Guskey, 2000; Loucks-Horsley et al., 2010; Villegas-Reimers, 2003). The goals need to be recognized and their

CONTACT Anja Kranjc Horvat kranjchorvat@uni-potsdam.de Physics Education Research, University of Potsdam, Karl-Liebknecht-Straße 24/25, 14476, Potsdam-Golm, Germany.

importance evaluated by all the stakeholders of the respective professional development (Guskey, 2000). Well-defined goals are crucial for the selection of both the appropriate professional development strategies (Loucks-Horsley et al., 2010) and the evaluation process itself (Guskey, 2000).

Therefore, our study investigated the goals of professional development programs (PDPs) at a particle physics laboratory from the perspective of four different stakeholder groups. The diversity of the stakeholder groups provided a broad overview and high heterogeneity of the emerging ideas. As such, our study is the first of its kind to compare the opinions of different groups of stakeholders on the goals of PDPs and their relative importance hierarchy.

State of the research

Building upon the conceptualization by Shulman (1986), the multifarious profession-specific knowledge that teachers possess can be described with three major knowledge categories: content knowledge (CK), pedagogical knowledge (PK), and pedagogical content knowledge (PCK) (e.g., Abell, 2007; Chan & Hume, 2019; Park & Oliver, 2008). First, content knowledge (CK) embodies the syntactic and substantive knowledge of the content selected for teaching (Abell, 2007). Riese et al. (2015), Enkrott et al. (2018), and Kulgemeyer and Riese (2018) define three steps of CK: school knowledge, university knowledge, and deeper school knowledge. Here, school knowledge and university knowledge strictly follow the high-school and the university curricula, respectively. Meanwhile, deeper school knowledge acts as a bridge between the two. Second, pedagogical knowledge (PK) encompasses general knowledge of learners and learning situations (Grossman, 1990). PK can be categorized into the knowledge of learning theories, knowledge of instructional principles, and knowledge of managing heterogeneous groups (e.g., Guerriero, 2017; Richardson, 1996). Last, pedagogical content knowledge (PCK) describes the knowledge of teaching the particular content (e.g., Abell, 2007; Park & Oliver, 2008; Shulman, 1986). Park and Oliver (2008) identified the following dimensions of PCK: knowledge of science curriculum, knowledge of instructional strategies for teaching science, knowledge of assessment, knowledge of students' understanding, teacher efficacy, and orientation in teaching in science.

Teachers' CK, PK, and PCK evolve through teaching experiences, professional sharing, and professional development opportunities (Schneider, 2019). Therefore, it is crucial to provide in-service teachers with various effective learning opportunities, including professional development programs (PDPs) to further develop their knowledge and skills in order to prepare them for new learning situations (Banilower et al., 2007; Borko, 2004; Greene et al., 2013; Hewson, 2007; Luft & Hewson, 2014; OECD, 2019; Pena-Lopez, 2009).

Effective professional development is defined by Darling-Hammond et al. (2017) as "structured professional learning that results in changes in teacher practices and improvements in student outcomes" (p. 2). However, the effectiveness of professional development opportunities does not rely on a single key feature, nor are the features of professional development independent. For example, short interventions are often scrutinized in the literature (e.g., Liu & Phelps, 2020; Smith & Gillespie, 2007) for being less effective than extensive programs. Indeed, Banilower et al. (2007) found a positive correlation between PDPs' length and their effectiveness, although the overall effects were small to moderate.

However, the PDP content and activities might have a greater influence on the effectiveness of a PDP than its length (Lipowsky & Rzejak, 2015; Wiener et al., 2018). Indeed, several studies linked the increase in teachers' knowledge and skills to how a PDP focuses on content knowledge (Birman et al., 2000; Darling-Hammond et al., 2017; Desimone, 2009; Wilson, 2013). Here, defining concrete goals is another key to effective PDPs (Guskey, 2000; Loucks-Horsley et al., 2010; Villegas-Reimers, 2003). Well-defined goals that are aligned with teachers' personal learning goals increase the likelihood of teachers becoming more ambitious and successful while learning (Coburn, 2004; Penuel et al., 2007; Zepeda, 2013). Furthermore, teachers' involvement in the development process increases their ownership, which leads to better implementation of PDP outcomes in their daily practice (Loucks-Horsley et al., 2010; Van Den Akker, 1999).

Different stakeholders can hold different perspectives on characteristics of effective PDPs, which can lead to disparities between theory and practice, expectations, and realizations (Park Rogers et al., 2007). An example of these disparities is visible in PDPs following top-down initiatives. Here, the PDP developers sometimes fail to consider the teachers and their personal learning goals; therefore, the teachers cannot implement the changes in the expected way (Collinson et al., 2009). Hence, finding a consensus between different stakeholders is an important step toward building effective PDPs (Brekelmans et al., 2013; Kennedy, 2007; O'Gorman & Drudy, 2011; Siko & Hess, 2014).

Learning goals for PDPs

Acknowledging teachers' goals for participating in professional development programs can influence their conceptual change process (Ebert & Crippen, 2010). These goals generally overlap with enhancing individual knowledge categories, namely PK, CK, and PCK, or their dimensions, as described above. However, the goals of PDPs generally stem from one of the four most commonly identified stakeholder groups, namely teachers, education researchers, research scientists, or policymakers. In the following paragraphs, we will give a literature overview of the goals of each stakeholder group.

First, we present the results of several studies on the learning goals of teachers participating in professional development (e.g., Anderson & Mitchener, 1994; Borko, 2004; Louws et al., 2018; Park Rogers et al., 2010; Van Duzor, 2012). Here, teachers identified three learning goals for their professional development: (1) to enhance their CK (Anderson & Mitchener, 1994; Louws et al., 2018; Owens et al., 2018; Park Rogers et al., 2010; Van Duzor, 2012), (2) to enhance their PK (Louws et al., 2018; Park Rogers et al., 2010), and (3) to enhance their PCK. In the latter, teachers specified various components of PCK. In particular, teachers aimed to learn about and gain experience in innovative instructional strategies and representations (Louws et al., 2018; Owens et al., 2018; Park Rogers et al., 2007, 2010) and increase their knowledge of the curricula (Owens et al., 2018; Van Duzor, 2012). Several studies also reported other goals that are important to teachers, such as improving their ability to use multimedia in their classrooms (Louws et al., 2018; Owens et al., 2018). Additionally, Owens et al. (2018) showed that teachers found it important to learn how to connect different fields of science, technology, and mathematics in their teaching and how to address gender inequalities.

Second, we present the goals of PDPs as perceived by research scientists who are involved in such programs but have no qualifications in education. The research scientists reported that their goals are for teachers to enhance their CK (Drayton & Falk, 2006; Gentsch, 1999; Schuster & Carlsen, 2009; Taylor et al., 2008) and to increase their knowledge of the curricula (Schuster & Carlsen, 2009). Moreover, Gentsch (1999) found that research scientists value the goal of enhancing teachers' CK more than enhancing their PK or PCK. However, it has been found that research scientists often struggle to understand what is appropriate for different grade levels (Loucks-Horsley et al., 2010). Furthermore, many have problems understanding and balancing the principles of adult learning and recent development in scientific research, which calls for cooperation with professionals in education (Loucks-Horsley et al., 2010).

Third, the opinions of education researchers regarding learning goals of PDPs have not been investigated in previous studies. Therefore, recommendations from the existing literature on PDPs were explored to approximate their opinions. As a whole, the studies recommend the following PDP goals as perceived by education researchers: enhancing CK (Borko, 2004; Desimone, 2009; Garet et al., 2001; Park Rogers et al., 2010; Smith & Gillespie, 2007), enhancing PK (Smith & Gillespie, 2007), and enhancing PCK (Borko, 2004; Desimone, 2009; Garet et al., 2001; Smith & Gillespie, 2007). Within the latter, Borko (2004) further distinguished between enriching teachers' knowledge of curricula and enhancing their knowledge of innovative instructional strategies. Furthermore, Van Driel and Berry (2012) showed that teachers should not only receive input on subject-specific instructional strategies but also gain opportunities to enact them in classroom settings. However, studies by Astor-Jack et al. (2007) and Luft and Hewson (2014) have found that the goals for professional development in science can vary.

Lastly, the goals of PDPs stated in various national policies are tightly connected to the curricula (e.g., Cormas & Barufaldi, 2011; Ebert & Crippen, 2010). However, policies in professional development are often stated ambiguously (Hardy et al., 2010; Kennedy, 2015) and its subsequent goals are also defined rather vaguely. The most common learning goals in policies include strengthening teachers' competencies (Hardy et al., 2010), keeping teachers in contact with new curricula and policies (Loeb et al., 2009), and improving teachers' understanding of high-quality academic research (Department for Education, 2016). Indeed, the literature rarely reports more specific learning goals, such as enhancing CK (Donnelly & Argyle, 2011). On this point, Hardy et al. (2010) found that unclear and sometimes contradictory policies can stem from the need for policymakers to respond to the requests of multiple PDP stakeholders.

The above-discussed stakeholders' goals do not include all dimensions of professional knowledge that have been defined in the literature (e.g., Abell, 2007; Grossman, 1990; Park & Oliver, 2008). Furthermore, different stakeholders have different expectations and competing values regarding teachers' professional development (Cervero & Daley, 2011; Teitel, 2001). However, several studies have found that collaboration between stakeholders in the design phase of a PDP can lead to better professional development (Brekelmans et al., 2013; Kennedy, 2007; O'Gorman & Drudy, 2011; Siko & Hess, 2014).

Research aim

As showcased above, studies have generally examined the expectations of a single stakeholder group regarding the goals of PDPs. We found a paucity of research dedicated to comparing the different groups of stakeholders on this topic. This is relevant as several studies (Brekelmans et al., 2013; Kennedy, 2007; O'Gorman & Drudy, 2011; Siko & Hess, 2014) have already recognized the importance of including multiple stakeholders in various fields in the design phase of PDPs. We hypothesize that incorporating stakeholders with different backgrounds into the research will result in a more specific hierarchy of goals. As such, it may be beneficial to any further research in this field. Indeed, a richer view of the research subject can be achieved by including different groups of stakeholders. Furthermore, knowing the positions of the different stakeholders regarding PDPs is crucial for establishing effective professional development practices (Brekelmans et al., 2013; Kennedy, 2007; O'Gorman & Drudy, 2011; Park Rogers et al., 2007; Siko & Hess, 2014).

The following two research questions guided our exploration of the hypothesis:

(1) Which learning goals of professional development programs (PDPs) for in-service high-school science teachers at science research institutions are perceived as the most important by the relevant stakeholders?
(2) What differences and similarities between the expectations of different groups of stakeholders regarding the learning goals of PDPs for in-service high-school science teachers at science research institutions can be identified?

Methodology

Answering the research questions required a robust research method. For this purpose, we conducted a conventional Delphi study. A Delphi study is an iterative method used to determine the ideas and judgments of experts on a particular topic (Osborne et al., 2003). We conducted this Delphi study in the context of PDPs at the European Organization for Nuclear Research (CERN), the largest particle physics laboratory, located in Geneva, Switzerland. Our study results will play an essential role in the evaluation of the programs.

Typically, a Delphi study includes two or more rounds of questionnaires with interspersed feedback (Clayton, 2006; Goldstein, 1975; Gupta & Clarke, 1996; Hsu & Sandford, 2007; Linstone & Turoff, 1975; Rowe & Wright, 1999). This design allows the participating experts to remain both geographically separated (Clayton, 2006; Enzer, 1975; Hasson et al., 2000; Osborne et al., 2003) and anonymous (Hasson et al., 2000; McMillan et al., 2016; Osborne et al., 2001; Rowe & Wright, 1999). Anonymity plays a significant role in reducing the social influence of individuals (Bolger & Wright, 2011) by ensuring that everyone's voice counts the same. As such, a Delphi study allows for a bigger group of experts from various fields and different parts of the world. However, the quality of a Delphi study depends mostly on the experiential knowledge of the selected expert group (Baker et al., 2006; Powell, 2003; Rowe & Wright, 1999). Therefore, their expertise needs to be carefully defined to fit the scope of the research.

A conventional Delphi study generally begins with an open or semi-open questionnaire (Clayton, 2006; Custer et al., 1999; Enzer, 1975; Hasson et al., 2000; Linstone & Turoff, 1975; Osborne et al., 2001; Reeves & Jauch, 1987; Rowe & Wright, 1999). This qualitative approach generates ideas and suggestions from the participating experts that are later used as the basis for subsequent questionnaires (Rowe & Wright, 1999). The participants in the study are asked in the subsequent rounds to make a judgment on the themes that emerged from the first round of the questionnaire (Linstone & Turoff, 1975) and the literature review (Hasson et al., 2000).

Between the rounds, the experts receive feedback from the researchers based on the analysis of the previous round. The feedback allows the experts to revise their judgments in light of others' judgments and build on the data throughout all the rounds (Brady, 2015; Clayton, 2006; Gupta & Clarke, 1996; Hsu & Sandford, 2007; Linstone & Turoff, 1975). Typically, a conventional Delphi design would conclude after three rounds of questionnaires (Hasson et al., 2000; Linstone & Turoff, 1975). This study also followed the three-round design. Each round is described in detail in the sections below.

The study's width and the number of participants depend significantly on the researchers' definition of an expert (Baker et al., 2006). Typically, the key characteristics of an expert are relevant knowledge, experience, and policy influence or position within key organizations (Baker et al., 2006; Osborne et al., 2001). The desired characteristics need to be determined before the experts are selected to ensure that the desired expertise is represented in the group (Okoli & Pawlowski, 2004).

However, a Delphi study with a heterogeneous group of participants will not necessarily illuminate any differences between the groups. Therefore, Jones (1975) suggests a panel structure for the participating stakeholders. Panels represent homogeneous groups of stakeholders with approximately equal qualifications. A panel structure helps provide a fairer representation of views and thus reduces the influence of more influential individuals (Bolger & Wright, 2011). In this case, most studies usually have between 10 and 25 participants per panel (Clayton, 2006; Okoli & Pawlowski, 2004; Osborne et al., 2003, 2001). Indeed, Hasson and Keeney (2011) showed that the reliability of a study improves in direct relation to the number of participants. However, very few new ideas are generated within a homogeneous group of more than 30 well-chosen individuals (Delbecq et al., 1975).

Participants

In our study, stakeholders were invited to form four distinct stakeholder groups: (a) physics education researchers, (b) representatives of the governments of different countries with experience in particle physics policies (hereafter referred to as government representatives), (c) research scientists in the fields of natural science and informational technology, and (d) high-school science teachers. All stakeholder groups and the respective qualifications of the stakeholders are described in Table 1.

The experts were selected using several techniques. First, the physics education researchers with experience in PDPs were identified using the snowballing technique. Their qualifications were assessed using predefined criteria (described in Table 1). Next,

Table 1. The table represents the four stakeholder groups and the qualifications of the respective stakeholders.

Stakeholder group	Qualifications
Physics education researchers	Experience in the research of PDPs or modern physics education.
Government representatives	Extensive knowledge of education programs at their respective institutions; experience with education policies in their respective home countries.
Research scientists	Researchers in natural sciences or informational technology; involved with the organization of national PDPs.
Teachers	High-school science teachers; participated in a PDP at CERN in the past or applied to participate in the future.

the government representatives were selected using a nomination technique. They were nominated based on their experience with the educational programs and systems of their institutions and countries, respectively. The group of research scientists was comprised of scientists in the fields of both natural sciences and informational technology. The participating scientists had been actively involved in the organization and facilitation of PDPs at CERN in their respective national languages. Last, the in-service teachers from various countries and educational systems worldwide were contacted either shortly before or after they had participated in their respective national or international PDP at CERN.

Following guidelines from the literature, more than ten experts participated in each stakeholder group (Clayton, 2006; Hasson & Keeney, 2011; Okoli & Pawlowski, 2004; Osborne et al., 2003, 2001), as seen in Table 2. Before the start of the Delphi study, the experts received an information package that included details on the study such as its purpose, length, the different stakeholder groups, and a brief overview of the first round.

Conduct of the study

In a conventional three-round Delphi study, the results of one round affect the methodology of the subsequent round. Therefore, the results of the first and second rounds are presented alongside their respective data collection descriptions. This form of presentation provides a better understanding of the methodology and the final results. The latter will be presented afterward.

In this Delphi study, we aimed to generate a variety of goals by eliciting free-formed ideas from the experts with the first-round open-ended question. Therefore, the first-round question was phrased more broadly than the research questions. The broad phrasing allowed the experts to include goals that might not have emerged from our literature review. The question was pretested through semi-structured interviews with three members of the stakeholder groups to ensure it was neither ambiguous nor leading. The updated question was again pretested with five physics education researchers to ensure that it was understandable by an international group of stakeholders. Finally, the pretests resulted in the following open-ended question:

> What are the goals of professional development programs at CERN and similar large research institutions?

Round 1: data collection and analysis

The first-round open-ended question was sent to the selected stakeholders with a one-month deadline for completion. The data from the first round were analyzed using the

Table 2. Number of stakeholders that populated each of the stakeholder groups in the three rounds.

Stakeholder group	1st Round	2nd Round	3rd Round
Physics education researchers	28	31	32
Government representatives	16	11	12
Research scientists	18	14	11
Teachers	19	45	43
Total	**81**	**101**	**98**

inductive thematic network analysis, as introduced by Attride-Stirling (2001). In the inductive approach, the patterns identified in the first 10 to 15 responses defined the majority of the themes, as suggested by Braun and Clarke (2006). The emergent themes connected into a web-like hierarchical network, with the themes clustered into bigger categories, as indicated by Attride-Stirling (2001).

An inter-rater analysis of 10% of the responses was conducted to improve the validity of the first-round analysis. The initial inter-rater agreement was 75%, which rose to 100% after the discussion. Additionally, the themes were compared to the results from the literature to identify differences. The results of the analysis were summarized in the feedback package for the experts. Within the package, all of the themes were briefly described and examples of the responses for each theme were added.

Round 1: results

The inductive thematic network analysis of the first-round questionnaire resulted in seven themes that showcase the various goals of PDPs. All of the emerging themes from the first round of our study had already been identified in previous studies (Borko, 2004; Hardy et al., 2010; Louws et al., 2018; Owens et al., 2018; Park Rogers et al., 2010), though not necessarily by all stakeholders. The themes were clustered into the three categories of professional knowledge: PK, CK, and PCK (Abell, 2007; Chan & Hume, 2019; Park & Oliver, 2008). The themes under PK and PCK correspond to the respective PK and PCK dimensions that were discussed above. The themes "Enhance knowledge of concepts and models" and "Learn about connections between different fields of science" match the definition of CK by Abell (2007) and Woitkowski & Borowski (2017). Based on the model of CK by Enkrott et al. (2018), the theme "enhance knowledge of solving equations," was added for the second-round questionnaire. The resulting eight themes and their respective categories are presented in Table 3 with respective examples.

The themes that emerged from the first-round analysis provided the basis for the second-round questionnaire. This questionnaire aimed to investigate the perceived importance of the eight themes. The online questionnaire was pretested by nine physics education researchers and two high-school science teachers. The pretest ensured that the questionnaire was understandable and could be completed within the desired time frame. None of the researchers and teachers in this pretest later participated in the Delphi study.

Round 2: data collection and analysis

In the second-round questionnaire, the stakeholders were asked to rate the importance of the themes from the first round. The rating was done on a 6-point Likert-type scale of importance, ranging from 1 ("very unimportant") to 6 ("very important"). An even number of points on the scale were chosen to avoid a neutral option and thus force the participants to choose a side, as proposed by Harvey and Holmes (2012) and Turoff (1975). The rating scale was introduced on the first page of the questionnaire, as seen in Clayton (2006).

As suggested by Boone and Boone (2012), the Likert-type questionnaire responses were descriptively analyzed. The central tendency and variability of the responses were assessed using medians and frequency analysis, respectively. The themes were ranked based on the percentage of stakeholders who agreed to rate the themes higher than "rather unimportant." The differences between the themes that were adjacent in the ranking scale were assessed

Table 3. Overview of the eight themes that emerged from the analysis of the first round of the Delphi study. In the first column, the knowledge category of each theme is denoted. Here, CK stands for content knowledge, PK for pedagogical knowledge, and PCK for pedagogical content knowledge. The last column presents examples of the responses that were included in the creation of the respective theme.

Category	Theme	Example
CK	Enhance knowledge of concepts and models	*Making clear how much science relies on modeling.*
CK	Enhance knowledge of solving equations	*Solving equations to describe physics phenomena*. (Based on Enkrott et al. [2018])
CK	Learn about connections between different fields of science	*Improvement of the teachers' knowledge (. . .) in connection to other fields.*
PK	Enhance general knowledge of learners and learning situations	*To keep teachers up to date with current research in pedagogy.*
PK	Learn to address gender inequalities	*Provide teachers with (. . .) ways to include gender diversity topics in their classrooms.*
PCK	Learn about new subject-specific instructional strategies	*Teachers need to learn how to teach physics.*
PCK	Gain experience in new subject-specific instructional strategies	*To personally experience new or alternative teaching approaches.*
PCK	Enhance knowledge of curriculum	*Increase the pedagogical content knowledge on how to include particle physics in the classroom.*

using the Mann-Whitney-Wilcoxon test. The differences between the stakeholder groups per theme were analyzed using the Kruskal-Wallis test. Furthermore, the stakeholder groups were compared pairwise per theme using the pairwise Wilcoxon signed-rank test.

Additionally, the stakeholders were encouraged to add comments to justify their ratings, change the wording of the themes, and suggest any merging, splitting, or adding of themes. The comments by the stakeholders were analyzed using thematic analysis. The results of the analysis helped ensure that the themes presented a valid and complete representation of their ideas and thus reduce the researchers' bias (Brady, 2015; Osborne et al., 2001). Indeed, the patterns arising from the comments influenced the phrasing of two themes for the third-round questionnaire and the creation of two additional themes.

The results of the analysis were summarized in the feedback package for the stakeholders. The feedback included both the ratings of all stakeholders and the stakeholder group ratings. The medians and frequencies were reported together with the Likert stacked bar chart for better visualization. Additionally, all statistically relevant differences between the stakeholder groups were summarized. Finally, the relevant comments were added. The feedback package was sent to the stakeholders together with an invitation to participate in the third-round questionnaire.

Round 2: results

The variability assessment of the second-round questionnaire showed a very prominent ceiling effect. As shown in Figure 1, all themes were rated "slightly important," "rather important," or "very important" by at least 64% of the stakeholders. Only the theme of "enhance knowledge of solving equations" was considered to be "very important" by less than a third of the stakeholders.

Additionally, the analysis of the central measure is presented in Table 4. Here, the theme of "enhance knowledge of solving equations" had a median of 4 ("slightly important"), the theme of "enhance knowledge of curricula" a median of 6 ("very important"), and other themes a median of 5 ("rather important").

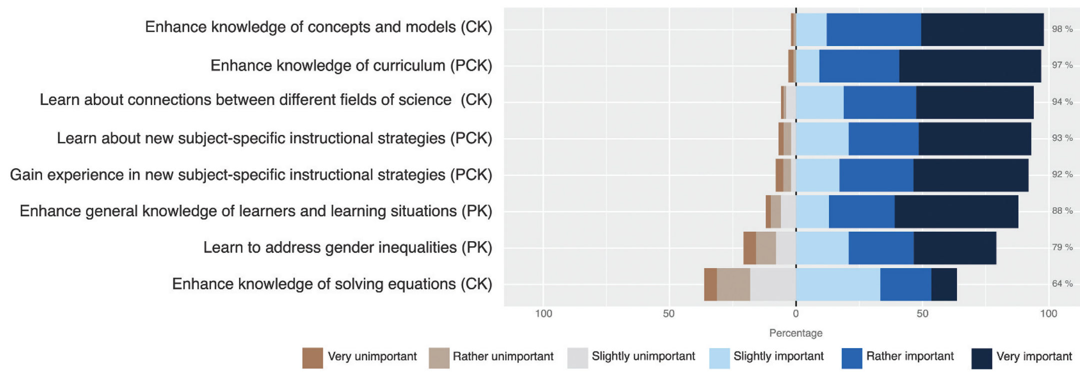

Figure 1. Frequency analysis of the responses to the second-round Likert-type questionnaire. The Likert stacked bar chart represents the combined responses of all stakeholder groups. The themes are sorted by the percentage of stakeholders who rated them at least "slightly important" (blue). Indeed, the ceiling effect is very apparent, as all themes were rated this way by more than 64% of the stakeholders.

Table 4. Results of the second-round rating of all stakeholders. The second column shows the percentage of stakeholders that rated the theme positively, namely "slightly important," "rather important," or "very important." More specifically, the third column only shows the percentage of stakeholders who rated themes as "very important." The last column shows the median for each theme. Here, 4 stands for "slightly important," 5 for "rather important," and 6 for "very important."

Theme	% Important	% Very important	Median
CK: Enhance knowledge of concepts and models	98	48.5	5
PCK: Enhance knowledge of curriculum	97	56.1	6
CK: Learn about connections between different fields of science	94	46.5	5
PCK: Learn about new subject-specific instructional strategies	93	44.6	5
PCK: Gain experience in new subject-specific instructional strategies	92	45.5	5
PK: Enhance general knowledge of learners and learning situations	88	49.0	5
PK: Learn to address gender inequalities	79	32.7	5
CK: Enhance knowledge of solving equations	64	10.1	4

The Mann-Whitney-Wilcoxon test between pairs of adjacent-ranked themes showed significant differences ($p < .05$) between the last three themes in Figure 1. "Learn to address gender inequalities" was rated significantly lower than "learn about new subject-specific instructional strategies" ($W = 6070$, $p = .01$). Additionally, "enhance knowledge of solving equations" was rated significantly lower than "learn to address gender inequalities" ($W = 6579$, $p < .001$). The ranking of the rest of the themes was not significantly different ($p > .05$).

An inductive thematic analysis of the 51 comments by the stakeholders revealed several patterns. The comments were mostly justifications for the ratings (i.e., the stakeholders would explain why something was more or less important). Some patterns showed the need for rephrasing a specific theme, for instance:

> A physics education researcher on learning to address gender inequalities: "This is a crucial part of the program, and organizers should absolutely consider adding ways to address racial inequality."

> A government representative on learning about connections between different fields of science: "Mainly, they should understand that the particle physics research is directly related with advanced technology, including electronics, advanced materials, informatics, computing, controls, etc."

In this case, both themes were rephrased to include updates from the comments. Indeed, the theme of gender inequality now also included other inequalities. At the same time, the theme of connecting different fields of science included a connection to technology. Additionally, several comments called for adding two new themes to be added:

> Government representative: "They should learn about different ways they could introduce the acquired knowledge into their classrooms and then test that knowledge."

> Teacher: "The use of ICT (information and communications technology) is unavoidable. It is good to learn how to use it better in teaching."

Similar patterns already appeared in the first round. However, they were included in the theme of learning about subject-specific instructional strategies. As the patterns repeated in the comments of the second round, two new themes were created: (1) enhancing the knowledge of the use of multimedia in instruction and (2) enhancing the knowledge of

Table 5. This table presents all of the themes in the final round of the questionnaire and their descriptions. The third column shows whether the theme stems from the first-round open-ended question, the literature, or the comments in the second round. The stars mark themes that were updated as a result of the analysis of the second-round comments.

Theme	Short description	Origin
CK: Enhance knowledge of concepts and models	Teachers learn relevant concepts and scientific models	Round 1
CK: Enhance knowledge of solving equations	Teachers learn various relevant mathematical equations and how to solve them	Literature
CK: Learn about connections between different fields of science and technology	Teachers learn how to connect different fields of science (*and technology) in their classrooms	Round 1*
PK: Enhance general knowledge of learners and learning situations	Teachers enhance their knowledge on creating effective and welcoming learning environments	Round 1
PK: Learn to address inequalities	Teachers learn how to address gender (*and other) inequalities in science and everyday life	Round 1*
PCK: Enhance knowledge of curriculum	Teachers learn how to connect the newly acquired knowledge to their existing curriculum	Round 1
PCK: Learn about subject-specific instructional strategies	Teachers learn different methods of presenting their respective subject to their students	Round 1
PCK: Gain experience in new subject-specific instructional strategies	Teachers experience and try out different methods of presenting their respective subject to their students	Round 1
PCK: Knowledge of assessments	Teachers learn about different ways of formative and summative knowledge assessments	Round 2
PCK: Enhance knowledge of the use of multimedia in instruction	Teachers learn how to use multimedia to better support their teaching	Round 2

assessments, which are both categories within PCK. The updated list of themes after the second-round questionnaire analysis is presented in Table 5.

Round 3: data collection and analysis

The third-round questionnaire was designed to evaluate the hierarchy of the themes presented in the second round. Because rating did not prove to be sufficient in revealing a hierarchy, the design of the third-round questionnaire included ranking tasks to force the differentiation. The participants were asked to rank the themes—described in Table 5—by their perceived importance, with the most important themes at the top. Again, a comment section was added to allow for any further comments on phrasing and ranking.

The ranking analysis was comprised of several steps. First, the overall ranking and the rankings of the individual stakeholder groups were constructed based on the medians and the interquartile ranges of the themes. The overall ranking was assessed for significance using the Mann-Whitney-Wilcoxon test. Here, the differences between 44% of the adjacently ranked themes (e.g., themes in third and fourth place) were not significant ($p > .05$). All pairs of themes with a significant difference were in the center of the ranking. The top two and bottom three themes were statistically undistinguishable ($p > .05$).

Therefore, the themes were grouped into three groups based on the positions of the medians of the themes relative to the quartiles of the overall ranking scale. As seen in Figure 2, the themes were grouped into (1) high importance themes, with medians falling into the first quartile; (2) medium importance themes, with medians falling into the interquartile range; and (3) low importance themes, with medians falling into the fourth quartile of the overall ranking. The differences between the ranked groups were assessed for significance using the Mann-Whitney-Wilcoxon test.

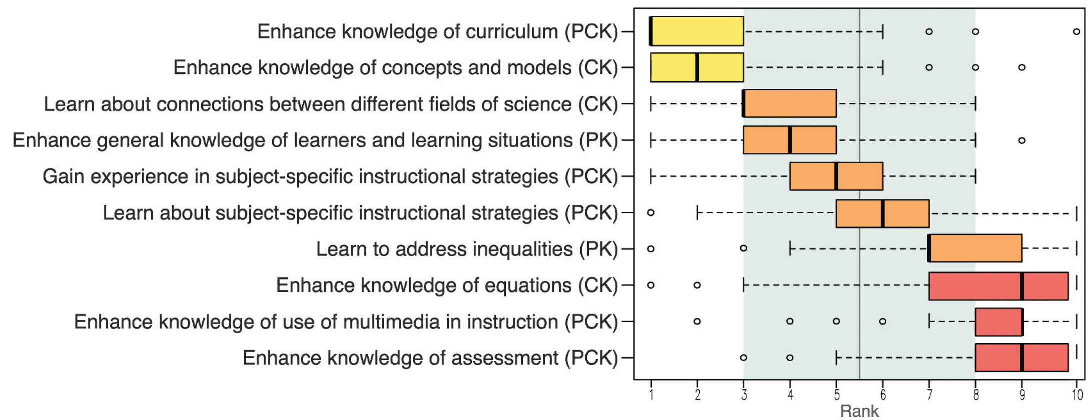

Figure 2. The boxplot shows the overall rankings of the ten learning goals and objectives. The shaded area on the graph represents the interquartile range of the whole set of themes used to differentiate between the high, medium, and low importance themes. The yellow bars represent high importance themes, orange bars represent medium importance themes, and red bars represent low importance themes.

The rankings per stakeholder group were compared using Kendall's W coefficient of concordance to assess the level of overall agreement. Disagreements between the stakeholder groups on individual themes were assessed using the Kruskal-Wallis test. Any disagreements were additionally analyzed with the Dunn's test with Bonferroni adjustment, as recommended by Dinno (2015). Furthermore, differences between the stakeholder groups on individual themes were analyzed using the pairwise Wilcoxon signed-rank test.

Additionally, as in the previous rounds, all experts were encouraged to provide their justifications, possible comments on the wording of the themes, and possible additions to the list. Seven comments were received in total. The comments were thematically analyzed to assess any previously unidentified patterns.

Final results

Overall, ten themes emerged from this three-round Delphi study. The final themes and results of the analysis of the final round are shown in Table 6. Based on the medians and the interquartile ranges, the themes have been grouped into high, medium, and low importance, as visualized in Figure 2. The Mann-Whitney-Wilcoxon test resulted in significant differences between all three groups ($p < .001$). The themes ranked "highly important" were "enhance knowledge of concepts and models" and "enhance knowledge of the curricula." On the other side, the lowest ranking themes were "enhance knowledge of solving equations," "enhance knowledge of the use of multimedia in instruction," and "enhance the knowledge of assessments."

The analysis of the agreement between stakeholder groups using Kendall's W coefficient of concordance showed strong agreement between stakeholder groups on the overall ranking ($W = 0.90$, $p < .01$). Furthermore, the Kruskal-Wallis test showed that the stakeholder groups' ratings were not significantly different in nine out of ten themes ($p > .05$). The only significant difference between the stakeholder groups is on the theme "learn about connections between different fields of science and technology" ($\chi^2 = 12.82$, $p = .005$). Here, the Dunn's test with Bonferroni adjustment showed a difference between physics education researchers and teachers ($Z = -3.49$, $p = .003$). The teachers found this theme to be significantly more important than physics education researchers did; however, they were still within the same ranking group (both ranked it as medium importance).

Table 6. The final ranking of the ten themes with the median ranking for each theme. Lower medians represent higher rankings (e.g., 1 is the highest rank). The last two columns present the results of the Kruskal-Wallis test of comparing the stakeholder groups, specifically the χ^2 and p-values. Significance estimates are marked as follows: $^{n.s.}p > .05$, $^{*}p < .01$.

Theme	Median	Importance	Kruskal-Wallis	
			χ^2	p-value
PCK: Enhance knowledge of curriculum	1	High	4.74	.20 n.s.
CK: Enhance knowledge of concepts and models	2	High	6.45	.10 n.s.
CK: Learn about connections between different fields of science	3	Medium	12.82	< .01*
PK: Enhance general knowledge of learners and learning situations	4	Medium	2.68	.44 n.s.
PCK: Gain experience in new subject-specific instructional strategies	5	Medium	6.84	.08 n.s.
PCK: Learn about subject-specific instructional strategies	6	Medium	1.69	.64 n.s.
PK: Learn to address inequalities	7	Medium	5.00	.18 n.s.
CK: Enhance knowledge of solving equations	8.5	Low	6.72	.08 n.s.
PCK: Enhance knowledge of the use of multimedia in instruction	9	Low	7.48	.06 n.s.
PCK: Knowledge of assessments	9	Low	5.91	.12 n.s.

Additionally, we split the group of high-school science teachers into two subgroups: (1) teachers who have participated in a PDP at CERN in the past and (2) teachers who have applied to participate in the future. The pairwise Dunn's test with Bonferroni adjustment showed a significant difference between physics education researchers and future participants of the program ($Z = -3.27$, $p = .013$). Indeed, future participants perceived the goal of "learn about connections between different fields of science and technology" as more important than the other stakeholder groups. Other comparisons between the subgroups of teachers and the other stakeholder groups, including a comparison between the two subgroups of teachers, showed no significant differences ($p > .05$).

The number of comments in the third round dropped to seven comments in total. All of the comments were directly related to the rankings. The only pattern that appeared was the difficulty to distinguish between the different themes in the ranking manners:

> Physics education researcher: "It is difficult to rank the above items in a strict manner. In my opinion, most of them should be equally valued."

No other patterns appeared in more than one comment and no patterns reemerged from the first or the second round.

Summary and discussion

This international Delphi study investigated the importance of learning goals of PDPs as perceived by different groups of stakeholders. More than 100 stakeholders from 42 countries provided their opinion on the subject by participating in three rounds of questionnaires. In the first round, the participants offered their ideas regarding the learning goals of PDPs. In the subsequent two rounds, these ideas were additionally evaluated by all the stakeholders to achieve the final ranking of all the suggested learning goals. As such, the study provided a unique comparison of the participating stakeholder groups and an overview of their perceptions of the goals of PDPs. Below, we summarize and discuss the study and its outcomes. Finally, we give an overview of possible implications of the study on the fields of PDPs and science education research.

Learning goals of PDPs

The first research question asked: "Which learning goals of PDPs for in-service high-school science teachers at science research institutions are perceived as the most important by the stakeholders of these PDPs?" Here, the stakeholders in our study identified nine dimensions of PCK, CK, and PK as the learning goals of PDPs. Additionally, one dimension was added from the literature for completeness. The stakeholders recognized all ten learning goals as important by rating and ranking the goals in the second and the third round, respectively. Similarly, all of the identified learning goals in our study have been recognized by at least one stakeholder group in previous studies (e.g., Borko, 2004; Donnelly & Argyle, 2011; Louws et al., 2018; Schuster & Carlsen, 2009). Furthermore, the stakeholders made no comments calling for any changes or additions to any learning goals in the last round of the study. Therefore, we can conclude that all of the important learning goals of PDPs at particle physics research institutions have been identified and clearly defined.

The majority of the identified learning goals fall into the categories of PCK and CK, as conceptualized in several previous studies (e.g., Abell, 2007; Chan & Hume, 2019; Park & Oliver, 2008). Indeed, previous studies showed that the two categories are often connected. Rollnick (2017) suggested that teachers' CK is enhanced more when the PDP also works on increasing the relevant PCK. Conversely, the development of PCK also relies on the advancement of CK (Davidowitz & Potgieter, 2016; Woitkowski & Borowski, 2017). Furthermore, different stakeholders in previous studies listed increasing both CK and PCK as the learning goals of PDPs (e.g., Borko, 2004; Donnelly & Argyle, 2011; Louws et al., 2018; Schuster & Carlsen, 2009). However, previous studies did not discuss the reasons why the increases in CK and PCK are mentioned more often by their stakeholders.

In the comments received in our study, several stakeholders argued that the goals related to enhancing PK should be addressed in PDPs at institutions for pedagogical research. This notion supports the findings by Astor-Jack et al. (2007) and Luft and Hewson (2014). They observed that the goals of PDPs in science vary depending on the type of institution. Science research institutions with many research scientists tend to offer a higher level of expertise in CK. However, the stakeholders in our study listed more components of PCK than that of CK. Furthermore, both were valued very similarly in the final rankings by all the stakeholders. This result was puzzling since the study by Gentsch (1999) found that research scientists valued increases in CK more than increases in PK and PCK. In that study, the research scientists were detached from education research. However, all research scientists in our study were strongly involved in science education due to their facilitation of PDPs. Their experiences in education might have influenced their priorities, which can thus explain the shift toward recognizing PCK components as important. Overall, the strong representation of PCK goals in the final list of learning goals calls on PDPs at science research institutions to include education researchers in the design and facilitation of PDPs.

Differences and similarities between the different groups of stakeholders

The second research question asked: "Which differences and similarities between the expectations of different groups of stakeholders regarding the learning goals of PDPs for in-service high-school science teachers at science research institutions can be identified?" On this point, most of the previous studies had only focused on individual stakeholder groups. Because these studies had been done in various contexts, their results are not directly comparable. Our study aimed to close that gap in the literature. Indeed, the results show strong agreement between the different stakeholder groups on both the learning goals of PDPs and their relative importance. Only the ranking of one learning goal showed a significant difference between the groups of stakeholders. The rankings of the rest of the learning goals showed no significant differences.

The only learning goal with a significant difference between any pair of the stakeholder groups is "learn to connect different fields of science and technology." Teachers ranked this learning goal as slightly more important than the education researchers had done. Similarly, this learning goal was also only mentioned by the teachers in one previous study (Owens et al., 2018). In our study, the teachers were represented by both past participants and applicants for future participation in a PDP at CERN. By comparing these two subgroups of teachers with other stakeholder groups, we see that the only statistical difference is between the future participants and the education researchers. Here, it can be assumed that the future participants are less familiar with the PDP's content, namely particle physics. Therefore, they might find connecting particle physics to the curriculum and real-life experiences more difficult. Indeed, this effect of our study might be less prominent in PDPs with a focus on topics that are more aligned with teachers' daily teaching practices. Nonetheless, we can conclude that the learning goals are well-aligned with both teachers and other stakeholders. Thus, the goals can be used as a basis for designing effective PDPs (Coburn, 2004; Penuel et al., 2007; Zepeda, 2013).

Strengths and limitations

Overall, our unique international study fills the gap in the literature by comparing the opinions of four different stakeholder groups on the learning goals of PDPs at particle physics laboratories. The study comprised over 100 experts from 42 countries, including experts from various research institutions in Europe and North America. Previous studies have combined the opinions of various stakeholder groups when determining the characteristics and limitations of their respective PDPs (e.g., Brekelmans et al., 2013; Kennedy, 2007; O'Gorman & Drudy, 2011; Siko & Hess, 2014). However, to our knowledge, no previous study has compared the opinions of various stakeholder groups regarding the learning goals of PDPs. Moreover, ours is one of the few studies that looked into the importance hierarchy of these learning goals. Here, Guskey (2000) notes that the goals of PDPs must be "considered important by all those involved in the professional development process" (Guskey, 2000, p. 89). Our research shows that the different groups of stakeholders of PDPs at CERN strongly agree on the hierarchy of the most important learning goals. Therefore, we are confident that our results represent the complete hierarchy of the ten most important goals of PDPs at particle physics laboratories.

However, several limitations of the study need to be considered. First, the study was conducted in the context of PDPs at the European Organization for Nuclear Research (CERN), the largest particle physics laboratory, located in Geneva, Switzerland. The field of particle physics is narrow and different from other fields of physics and science in general. Therefore, the outcomes of our study might have been influenced by the context of particle physics itself. Indeed, as shown by Astor-Jack et al. (2007) and Luft and Hewson (2014), the goals of science PDPs can vary. However, the final list of learning goals is similar to that found in the existing literature. Therefore, our study's findings will allow subsequent studies to omit the first two rounds and assess the hierarchy of the learning goals from our study within their respective contexts.

Second, the study is limited by the participants' expertise, which is common in Delphi studies (Clayton, 2006; Powell, 2003). Indeed, the participating stakeholders are not necessarily representative samples of their groups. An example of this limitation is the teachers. Indeed, the teachers participating in this study comprised both past and future participants of PDPs at CERN. Therefore, they were self-selected and likely highly motivated. The teachers' motivation for the field of particle physics and the PDPs in this context gives them unique expertise, which is required by the Delphi study design. The same is not necessarily true for teachers that have not applied for a PDP at CERN. Teachers with no intention of applying to this PDP have not been included in the study as they had no experience and knowledge of its workings. However, the teachers applying to similar PDPs would likely have similar characteristics to the teachers included in our study.

Third, due to the study's internationality, the questionnaires were administered in English. As most of the experts were not native speakers, the questionnaires were pretested within an international group of researchers to reduce the language bias. Furthermore, the option "I do not understand the question" was added in the third round of the questionnaire. This option helped to identify any issues with the language or phrasing of the question. It is worth noting that none of the participants reported a lack of understanding of the question; thus, a potential language bias is likely of minor effect.

Implications

The importance assessment of the learning goals of PDPs is considered to be valuable not only for the design of effective PDPs but also for their evaluation (Guskey, 2000). Our study provided an importance hierarchy of the ten most important learning goals as perceived by all stakeholders. The majority of previous studies often only listed the learning goals of the respective PDPs (e.g., Anderson & Mitchener, 1994; Borko, 2004; Louws et al., 2018; Park Rogers et al., 2010; Van Duzor, 2012). Indeed, very few studies have compared their relative importance (e.g., Gentsch, 1999). While a generalization of the results of our study on all PDPs at science research institutions might be a hasty conclusion to make, this study provides a strong case for the learning goals of PDPs at particle physics research laboratories. Therefore, the hierarchy of the goals, combined with the strong agreement of all stakeholders, can serve as a starting point for planning and evaluation of effective PDPs at all similar research institutions. For example, the two most important learning goals that emerged from this study are to "enhance knowledge of concepts and models" and to "enhance knowledge of the curricula." Their high importance ranking calls for a stronger focus during PDPs on relevant concepts and their implementation in the curriculum. In contrast, "enhancing knowledge of solving equations" was ranked significantly lower, which corroborates the implication to focus on enhancing teachers' conceptual rather than procedural knowledge.

The outcomes of our study also provide a strong starting point for further studies. A better generalization of the results would be possible with similar studies done at science research laboratories in different fields of science. In that case, the perceived importance of the identified learning goals might differ, as shown in the studies by Astor-Jack et al. (2007) and Luft and Hewson (2014). Furthermore, similar studies could include new relevant groups of stakeholders. For instance, teachers who have not applied to a PDP could be included to assess the differences between them and those who have previously applied. Future studies could reduce the number of rounds of questionnaires and thus the necessary time and effort by starting from the learning goals defined by our presented study.

Overall, the generalization by expanding our study with the proposed studies would support policymakers in the field of PDPs in establishing more specific education policies that focus on enhancing both teachers' CK and PCK. Consequently, a higher focus on CK and PCK, as proposed by our study, calls for a stronger collaboration between research scientists and education researchers. In return, this collaborative effort could lead to more effective design, facilitation, and evaluation of PDPs at science research institutions.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Anja Kranjc Horvat http://orcid.org/0000-0002-2584-4715
Jeff Wiener http://orcid.org/0000-0001-5817-9996
Sascha Schmeling http://orcid.org/0000-0003-2454-1699
Andreas Borowski http://orcid.org/0000-0002-9502-0420

References

Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1105–1149). Lawrence Erlbaum Associates.

Anderson, R. D., & Mitchener, C. P. (1994). Research on science teacher education. In E. D. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 3–44). Macmillan Publishing Company.

Astor-Jack, T., McCallie, E., & Balcerzak, P. (2007). Academic and informal science education practitioner views about professional development in science education. *Science Teacher Education*, *91*(4), 604–628. https://doi.org/10.1002/sce.20205

Attride-Stirling, J. (2001). Thematic networks: An analytic tool for qualitative research. *Qualitative Research*, *1*(3), 385–405. https://doi.org/10.1177/146879410100100307

Baker, J. A., Lovell, K., & Harris, N. (2006). How expert are the experts? An exploration of the concept of 'expert' within Delphi panel techniques. *Nurse Researcher*, *14*(1), 59–70. https://doi.org/10.7748/nr2006.10.14.1.59.c6010

Banilower, E. R., Heck, D. J., & Weiss, I. R. (2007). Can professional development make the vision of the standards a reality? The impact of the national science foundation's local systemic change through teacher enhancement initiative. *Journal of Research in Science Teaching*, *44*(3), 375–395. https://doi.org/10.1002/tea.20145

Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. *Educational Leadership*, *57*(8), 28–33.

Bolger, F. M. I., & Wright, G. (2011). Improving the Delphi process: Lessons from social psychological research. *Technological Forecasting and Social Change*, *78*(9), 1500–1513. https://doi.org/10.1016/j.techfore.2011.07.007

Boone, H. N., & Boone, D. A. (2012). Analyzing Likert data. *Journal of Extension*, *50*(2), 1–5.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Brady, S. R. (2015). Utilising and adapting the Delphi method for use in qualitative research. *International Journal of Qualitative Methods*, *14*(5), 1–6. https://doi.org/10.1177/1609406915621381

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brekelmans, G. F., Poell, R., & Van Wijk, K. (2013). Factors influencing continuing professional development: A Delphi study among nursing experts. *European Journal of Training and Development*, *37*(3), 313–325. https://doi.org/10.1108/03090591311312769

Cervero, B. M., & Daley, B. J. (2011). Continuing professional education: Multiple stakeholders and agendas. In K. Rubenson (Ed.), *Adult learning and education* (pp. 140–145). Elsevier.

Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 3–76). Springer Singapore. https://doi.org/10.1007/978-981-13-5898-2_1

Clayton, M. J. (2006). Delphi: A technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology*, *17*(4), 373–386. https://doi.org/10.1080/0144341970170401

Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, *77*(3), 211–244. https://doi.org/10.1177/003804070407700302

Collinson, V., Kozina, E., Kate Lin, Y. H., Ling, L., Matheson, I., Newcombe, L., & Zogla, I. (2009). Professional development for teachers: A world of change. *European Journal of Teacher Education*, *32*(1), 3–19. https://doi.org/10.1080/02619760802553022

Corcoran, T. B. (1995). *Helping teachers teach well: Transforming professional development*. CPRE policy briefs. http://repository.upenn.edu/cpre_policybriefs/74

Cormas, P. C., & Barufaldi, J. P. (2011). The effective research-based characteristics of professional development of the national science foundation's GK-12 program. *Journal of Science Teacher Education*, *22*(3), 255–272. https://doi.org/10.1007/s10972-011-9228-1

Custer, R. L., Scarcella, J. A., & Stewart, B. R. (1999). The modified Delphi technique—A rotational modification. *Journal of Vocational and Technical Education*, *15*(2), 50–58. http://doi.org/10.21061/jcte.v15i2.702

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.

Davidowitz, B., & Potgieter, M. (2016). Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry. *International Journal of Science Education*, *38*(9), 1483–1503. https://doi.org/10.1080/09500693.2016.1196843

Delbecq, A. L., Van De Ven, A. H., & Gustafson, D. H. (1975). *Group techniques for program planning: A guide to nominal group and delphi processes*. Scott, Foresman and Company.

Department for Education. (2016). *Standard for teachers' professional development* (DFE-00167-2016). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/537031/160712_-_PD_Expert_Group_Guidance.pdf

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Dinno, A. (2015). Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *The Stata Journal: Promoting Communications on Statistics and Stata*, *15*(1), 292–300. https://doi.org/10.1177/1536867X1501500117

Donnelly, L. A., & Argyle, S. (2011). Teachers' willingness to adopt nature of science activities following a physical science professional development. *Journal of Science Teacher Education*, *22*(6), 475–490. https://doi.org/10.1007/s10972-011-9249-9

Drayton, B., & Falk, J. (2006). Dimensions that shape teacher–scientist collaborations for teacher enhancement. *Science Education*, *90*(4), 734–761. https://doi.org/10.1002/sce.20138

Ebert, E. K., & Crippen, K. J. (2010). Applying a cognitive-affective model of conceptual change to professional development. *Journal of Science Teacher Education*, *21*(3), 371–388. https://doi.org/10.1007/s10972-009-9183-2

Enkrott, P., Buschhüter, D., & Borowski, A. (2018). Modeling and development of professional content knowledge of pre-service physics teachers. In C. Maurer (Ed.), *Qualitätsvoller Chemie- und Physikunterricht- normative und empirische Dimensionen. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Regensburg 2017* (pp. 879–882). Universität Regensburg.

Enzer, S. (1975). Plastics and competing materials by 1985: A Delphi forecasting study. In H. A. Linstone, & M. Turoff (Eds.), *The Delphi method: Techniques and applications* (pp. 189–203). Addison-Wesley Educational Publishers Inc.

Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F., & Suk Yoon, K. (2001). What makes professional development effective? Results from a national sample of teachers. *American Education Research Journal*, *38*(4), 915–945. https://doi.org/10.3102/00028312038004915

Gentsch, K. L. H. (1999). Science education goals: Views from scientists and teachers. *Journal of Elementary Science Education*, *11*(1), 61–68. https://doi.org/10.1007/BF03173791

Goldstein, N. H. (1975). A Delphi on the future of the steel and ferroalloy industries. In H. A. Linstone, & M. Turoff (Eds.), *The Delphi method: Techniques and applications* (pp. 204–220). Addison-Wesley Educational Publishers Inc.

Greene, B. A., Lubin, I. A., Slater, J. L., & Walden, S. E. (2013). Mapping changes in science teachers' content knowledge: Concept maps and authentic professional development. *Journal of Science Education and Technology*, *22*(3), 287–299. https://doi.org/10.1007/s10956-012-9393-9

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. Teachers College Press, Teachers College, Columbia University.

Guerriero, S. (2017). Pedagogical knowledge: What it is and how it functions. In S. Guerriero (Ed.), *Pedagogical knowledge and the changing nature of the teaching profession* (pp. 99–118). OECD. https://doi.org/10.1787/9789264270695-6-en

Gupta, U. G., & Clarke, R. E. (1996). Theory and applications of the Delphi technique: A bibliography (1975–1994). *Technological Forecasting and Social Change*, *53*(2), 185–211. https://doi.org/10.1016/s0040-1625(96)00094-7

Guskey, T. R. (2000). *Evaluating professional development*. Corwin Press, Inc.

Hardy, I., Rönnerman, K., Moksnes Furu, E., Salo, P., & Forsman, L. (2010). Professional development policy and politics across international contexts: From mutuality to measurability? *Pedagogy, Culture & Society*, *18*(1), 81–92. https://doi.org/10.1080/14681360903556871

Harvey, N., & Holmes, C. A. (2012). Nominal group technique: An effective method for obtaining consensus. *International Journal of Nursing Practice*, *18*(2), 188–194. https://doi.org/10.1111/j.1440-172X.2012.02017.x

Hasson, F., & Keeney, S. (2011). Enhancing rigour in the Delphi technique research. *Technological Forecasting and Social Change*, *78*(9), 1695–1704. https://doi.org/10.1016/j.techfore.2011.04.005

Hasson, F., Keeney, S., & McKenna, H. P. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008–1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x

Hewson, P. W. (2007). Teacher professional development in science. In S. K. Abell, & N. G. Lederman (Eds.), *Handbook of research on science education* (Vol. 1, pp. 1179–1203). Routledge.

Hsu, -C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation*, *12*(10), 1–8. https://doi.org/10.7275/pdz9-th90

Jones, C. G. (1975). A Delphi evaluation of agreement between organisations. In H. A. Linstone, & M. Turoff (Eds.), *The Delphi method: Techniques and applications* (pp. 155–161). Addison-Wesley Educational Publishers Inc.

Kennedy, A. (2007). Continuing professional development (CPD) policy and the discourse of teacher professionalism in Scotland. *Research Papers in Education*, *22*(1), 95–111. https://doi.org/10.1080/02671520601152128

Kennedy, A. (2015). What do professional learning policies say about purposes of teacher education? *Asia-Pacific Journal of Teacher Education*, *43*(3), 183–194. https://doi.org/10.1080/1359866X.2014.940279

Kulgemeyer, C., & Riese, J. (2018). From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching*, *55*(10), 1393–1418. https://doi.org/10.1002/tea.21457

Kurnaz, M. A., & Çepni, S. (2012). An evaluation of changes to the Turkish high school physics curriculum. *International Education Studies*, *5*(5), 92. https://doi.org/10.5539/ies.v5n5p92

Linstone, H. A., & Turoff, M. (1975). Introduction. In H. A. Linstone, & M. Turoff (Eds.), *The Delphi method: Techniques and applications* (pp. 3–14). Addison-Wesley Educational Publishers Inc.

Lipowsky, F., & Rzejak, D. (2015). Key features of effective professional development programmes for teachers. *Ricercazione*, *7*(2), 27–51.

Liu, S., & Phelps, G. (2020). Does teacher learning last? Understanding how much teachers retain their knowledge after professional development. *Journal of Teacher Education*, *71*(5), 537–550. https://doi.org/10.1177/0022487119886290

Loeb, S., Miller, L. C., & Strunk, K. O. (2009). The state role in teacher professional development and education throughout teachers' careers. *Education Finance and Policy*, *4*(2), 212–228. https://doi.org/10.1162/edfp.2009.4.2.212

Loucks-Horsley, S., Stiles, K. E., Mundry, S., Love, N., & Hewson, P. W. (2010). *Designing professional development for teachers of science and mathematics* (3rd ed. Expanded ed.). Corwin Press.
Louws, M. L., Meirink, J. A., Van Veen, K., & Van Driel, J. H. (2018). Understanding teachers' professional learning goals from their current professional concerns. *Teachers and Teaching*, *24*(1), 63–80. https://doi.org/10.1080/13540602.2017.1383237
Luft, J. A., & Hewson, P. W. (2014). Research on professional development programs in science. In N. G. Lederman, & S. K. Abell (Eds.), *Handbook of research on science education* (Vol. 2, pp. 97–118). Routledge.
McMillan, S. S., King, M., & Tully, M. P. (2016). How to use the nominal group and Delphi techniques. *International Journal of Clinical Pharmacy*, *38*(3), 655–662. https://doi.org/10.1007/s11096-016-0257-x
O'Gorman, E., & Drudy, S. (2011). *Professional development for teachers working in special education/inclusion in mainstream schools: The views of teachers and other stakeholders* [Research report]. University College Dublin.
OECD. (2019). *TALIS 2018 results (Volume 1): Teachers and school leaders as lifelong learners*.
Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management*, *42*(1), 15–29. https://doi.org/10.1016/j.im.2003.11.002
Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "ideas-about-science" should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, *40*(7), 692–720. https://doi.org/10.1002/tea.10105
Osborne, J., Ratcliffe, M., Collins, S., Millar, R., & Duschl, R. (2001). *What should we teach about science? A Delphi study*. Evidence-based Practice in Science Education (EPSE).
Owens, D. C., Sadler, T. D., Murakami, C. D., & Tsai, C.-L. (2018). Teachers' views on and preferences for meeting their professional development needs in STEM. *School Science and Mathematics*, *118*(8), 370–384. https://doi.org/10.1111/ssm.12306
Park Rogers, M. A., Abell, S. K., Lannin, J., Wang, C., Musikul, K., Barker, D., & Dinkman, S. (2007). Effective professional development in science and mathematics education: Teachers' and facilitators' views. *International Journal of Science and Mathematics Education*, *5*(3), 507–532. https://doi.org/10.1007/s10763-006-9053-8
Park Rogers, M. A., Abell, S. K., Marra, R. M., Arbaugh, F., Hutchins, K. L., & Cole, J. S. (2010). Orientations to science teacher professional development: An exploratory study. *Journal of Science Teacher Education*, *21*(3), 309–328. https://doi.org/10.1007/s10972-009-9179-y
Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, *38*(3), 261–284. https://doi.org/10.1007/s11165-007-9049-6
Pena-Lopez, I. (2009). *Creating effective teaching and learning environments: First results from TALIS*. Organisation for Economic Cooperation and Development (OECD).
Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Education Research Journal*, *44*(4), 921–958. https://doi.org/10.3102/0002831207308221
Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing*, *41*(4), 376–382. https://doi.org/10.1046/j.1365-2648.2003.02537.x
Reeves, G., & Jauch, L. R. (1987). Curriculum development through Delphi. *Research in Higher Education*, *8*(2), 157–168. https://doi.org/10.1007/BF00992116
Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Ed.), *Handbook of research on teacher education* (2nd ed., pp. 102–119). Macmillan.
Riese, J., Kulgemeyer, C., Zander, S., Borowski, A., Fischer, H., Gramzow, Y., Reinhold, P., Schecker, H., & Tomczyszyn, E. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. In S. Blömeke, & O. Zlatkin-Troitschanskaia (Eds.), *Kompetenzen von Studierenden: 61. Beiheft der Zeitschrift für Pädagogik* (pp. 55–79). Beltz.
Rollnick, M. (2017). Learning about semi conductors for teaching—The role played by content knowledge in pedagogical content knowledge (PCK) development. *Research in Science Education*, *47*(4), 833–868. https://doi.org/10.1007/s11165-016-9530-1

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting*, *15*(4), 353–375. https://doi.org/10.1016/S0169-2070(99)00018-7

Schneider, R. M. (2019). Illustrating and developing science teachers' pedagogical content knowledge through learning study. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 165–184). Springer, Singapore.

Schuster, D. A., & Carlsen, W. S. (2009). Scientists' teaching orientations in the context of teacher professional development. *Science Education*, *93*(4), 635–655. https://doi.org/10.1002/sce.20310

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Siko, J. P., & Hess, A. N. (2014). Win-win professional development while meeting the needs of all stakeholders. *TechTrends*, *58*(6), 99–108. https://doi.org/10.1007/s11528-014-0809-7

Smith, C., & Gillespie, M. (2007). Research on professional development and teacher change: Implications for adult basic education. *Review of Adult Learning and Literacy*, *7*(7), 205–244.

Taylor, A. R., Jones, M. G., Broadwell, B., & Oppewal, T. (2008). Creativity, inquiry, or accountability? Scientists' and teachers' perceptions of science education. *Science Education*, *92*(6), 1058–1075. https://doi.org/10.1002/sce.20272

Teitel, L. (2001). An assessment framework for professional development schools: Going beyond the leap of faith. *Journal of Teacher Education*, *52*(1), 57–69. https://doi.org/10.1177/0022487101052001006

Turoff, M. (1975). The policy Delphi. In H. A. Linstone, & M. Turoff (Eds.), *The Delphi method: Techniques and applications* (pp. 80–96). Addison-Wesley Educational Publishers Inc.

Van Den Akker, J. (1999). Principles and methods of development research. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), *Design approaches and tools in education and training* (pp. 1–14). Springer.

Van Driel, J. H., & Berry, A. (2012). Teacher professional development focusing on pedagogical content knowledge. *Educational Researcher*, *41*(1), 26–28. https://doi.org/10.3102/0013189X11431010

Van Duzor, A. G. (2012). Evidence that teacher interactions with pedagogical contexts facilitate chemistry-content learning in K-8 professional development. *Journal of Science Teacher Education*, *23*(5), 481–502. https://doi.org/10.1007/s10972-012-9290-3

Villegas-Reimers, E. (2003). *Teacher professional development: An international review of the literature*. International Institute for Educational Planning.

Wallace, C. S., & Priestley, M. R. (2017). Secondary science teachers as curriculum makers: Mapping and designing Scotland's new curriculum for excellence: Secondary science teachers as curriculum makers. *Journal of Research in Science Teaching*, *54*(3), 324–349. https://doi.org/10.1002/tea.21346

Wiener, G. J., Schmeling, S. M., & Hopf, M. (2018). The technique of probing acceptance as a tool for teachers' professional development: A PCK study. *Journal of Research in Science Teaching*, *55*(6), 849–875. https://doi.org/10.1002/tea.21442

Wilson, S. M. (2013). Professional development for science teachers. *Science*, *340*(6130), 310–313. https://doi.org/10.1126/science.1230725

Woitkowski, D., & Borowski, A. (2017). Fachwissen im Lehramtsstudium Physik. In H. Fischler, & E. Sumfleth (Eds.), *Professionelle Kompetenz von Lehrkräften der Chemie und Physik* (pp. 57–74). Logos.

Zepeda, S. J. (2013). *Professional development: What works*. Taylor and Francis.