ENHANCING PRIMARY TEACHER EDUCATION: RESEARCH-BASED DEVELOPMENT OF A PEDAGOGICAL CONTENT KNOWLEDGE RUBRIC

B. Valente[1], P. Maurício[1], M.J. Silva[1], N. Melo[1], P. Sarreira[2]

[1]Escola Superior de Educação de Lisboa, Ci&DEI, Instituto Politécnico de Lisboa (PORTUGAL)
[2]Escola Superior de Educação de Lisboa, Instituto Politécnico de Lisboa (PORTUGAL)

Abstract

Pedagogical Content Knowledge (PCK), introduced by Shulman [1], is crucial for teaching quality and student outcomes [2]. Shulman [1], [3] highlighted PCK as the key difference between a subject matter expert and an effective teacher, requiring both deep content understanding and the ability to implement effective teaching strategies.

Since its introduction, PCK has been further reconceptualized and developed by several authors [4]. However, its elusive nature [5] makes it necessary to develop instruments that make it visible. Developing robust tools to measure PCK is essential for improving teacher education programs. This study details the development and validation of a rubric designed to measure the PCK of preservice primary teachers.

The development of the PCK rubric followed a systematic process involving several key steps. First, a comprehensive review of existing literature on PCK and related assessment tools was conducted. This review provided a foundation for identifying the key dimensions of PCK relevant to preservice primary teachers. Based on the literature review, the critical components of PCK were identified, including curricular knowledge, knowledge about students, knowledge about teaching strategies, and knowledge about assessment. For each dimension, specific performance criteria were established to capture the range of preservice teacher capabilities.

The rubric's validation involved several phases to ensure its reliability and validity. An initial version of the rubric was reviewed by a panel of education experts, whose feedback guided the refinement of criteria and descriptors. The revised rubric was subsequently pilot tested with a sample of preservice teachers, using two different instruments for assessment: i) CoRe (Content Representation) [6], which requires preservice teachers to select and justify key ideas essential for teaching a specific topic; and ii) lesson plans, specifically for a one-and-a-half-hour lesson on the respiratory system for third-grade students. The topic was chosen based on its familiarity to students, its abstract nature, and its inclusion in the Portuguese curriculum [7].

Inter-rater reliability was assessed by having multiple evaluators use the rubric to score the same set of teaching artifacts. The consistency of scores across different raters was analysed to ensure the rubric's reliability. The validation process indicated that the PCK rubric requires further refinement to serve as a valid tool for assessing PCK.

Keywords: Pedagogical Content Knowledge, Preservice teachers Education, Rubric.

1 INTRODUCTION

Understanding the knowledge base for teaching – the knowledge that a teacher needs for instruction – was the driving force behind the introduction of Pedagogical Content Knowledge (PCK) as one of its seven categories by [3]. For Shulman [3], PCK was that “special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding” and “represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests of learners, and presented for instruction” (p. 8).

Since [3] proposal, the construct has been developed, refined, empirically studied, and widely used in teacher training, with most studies taking place in mathematics and science education [8], [9]. One model widely adopted in science education was developed by [10]. For the authors of this model, PCK constitutes a distinct domain of knowledge comprising five components: 1) orientations to science

teaching; 2) knowledge and beliefs about curricula; 3) knowledge and beliefs about assessment; 4) knowledge and beliefs about students' understanding; and 5) knowledge and beliefs about instructional strategies. This model extends Shulman's conceptualization by emphasizing the inherent connection between knowledge and beliefs, and also by introducing the first three components, which were not initially considered.

Park and Oliver [11] refined the model proposed by [10], creating the well-known hexagon model. According to authors of the hexagon model, PCK is idiosyncratic, and its development arises from both knowledge acquisition and knowledge use (knowledge-in-action and knowledge-on-action). As [12] state, "it is unlikely that teachers acquire PCK first, and then enact it" (p. 278). Another component of this model is teacher efficacy, or "teachers' beliefs in their ability to affect student outcomes" (ibid.). Finally, this model highlights the importance of teachers' understanding of students for the development of PCK, particularly that "teachers' understanding of students' misconceptions appeared to be a primary factor that influenced the teachers' PCK" (ibid.).

While theoretical and empirical research has led to a proliferation of PCK conceptualizations [13], [5], in the field of science education, efforts to find common ground to foster the use of PCK and allow for the comparison of results have increased in recent years, with emphasis on two major events: the 1st and 2nd PCK summits held in 2012 and 2016, respectively [4] and [14]. The Refined Consensual Model (RCM) emerged as a major outcome. This model consists of five components: three related to knowledge and skills in 1) curricular saliency, 2) conceptual teaching strategies, and 3) students' understanding of science. Additionally, it incorporates the integration of PCK components with pedagogical reasoning. The RCM is a dynamic model in which researchers identified personal, collective, and enacted realms of PCK and their interrelations.

Given the complex nature of PCK, one of the main challenges lies in the lack of a clear definition, which makes it difficult to specify the construct for assessment and, subsequently, to determine appropriate assessment methods. Another challenge in assessing PCK is its tacit nature, as teachers are often unaware of the knowledge they possess [15]. Furthermore, teachers may be unable or unwilling to articulate their understanding of PCK because there is no language or structure that adequately captures this knowledge [16], [17]. These factors occasionally make it difficult for researchers to elicit PCK from teachers. Most importantly, PCK is specific to topic, context, and individual, making it challenging to develop measures that are applicable across different topics, contexts, and teachers [15], [18]. This complexity particularly contributes to the difficulty in establishing a normative framework for comparing PCK among individual teachers working in different classroom settings.

Despite the inherent challenges in defining and assessing PCK, efforts to develop reliable methods for its evaluation continue to be crucial. The complexity of PCK, coupled with its context-specific nature, has led researchers to explore various tools for assessment. One such tool is the use of scoring rubrics, which rests on the widely accepted assumption that PCK can be measured along a continuum from weak to strong [19].

This paper is part of a larger project (IPL/IDI&CA2023/PCK-PST/ESELx) aimed at examining the development of PCK during the initial training of primary teachers, with the goal of identifying the factors or conditions that contribute to this development. Specifically, we present the development of a scoring rubric to assess elementary preservice teachers' PCK.

2 METHODOLOGY

The development of the PCK rubric followed a structured, multi-step process. It began with a comprehensive review, by two of the authors, of the literature on PCK and relevant assessment tools, which informed the selection of instruments for data collection. Key factors, such as the specific PCK components to be evaluated, the timing of assessments throughout the master's program, and the suitability of the instruments, were carefully considered. This review also provided a foundation for identifying the essential PCK components for preservice primary teachers, ensuring that the rubric captured the most critical elements of PCK.

The next step, closely linked to the initial review, involved examining existing rubrics used to assess PCK through data collection tools similar to those selected earlier. This analysis, along with a detailed examination of the collected data, guided the creation of a new rubric tailored to the study's needs.

To ensure the rubric's reliability and validity, it underwent a thorough validation process. An initial version was reviewed by a panel of education experts, the authors, whose feedback led to the refinement of the

criteria and descriptors. The revised rubric was subsequently pilot tested with a sample of preservice teachers. Multiple evaluators used the rubric to score the same set of teaching artifacts, allowing for an analysis of inter-reliability between coders.

According to [20], inter-rater reliability (IRR) "provides a way of quantifying the degree of agreement between two or more coders who make independent ratings about the features of a set of subjects" (p. 1). Following [20] recommendations, several decisions were made to assess the inter-rater reliability of the rubric. First, we decided that a subset of preservice teachers would be rated by multiple coders, with the remainder rated by single coders. Second, the subjects rated by multiple coders were rated by the same set of coders (fully crossed design), allowing for systematic bias between coders to be assessed and controlled in the IRR estimate, in order to improve overall IRR estimates. Third, we conducted extensive training before coding subjects from the actual study.

Given that the data used in this rubric are ordinal, the intra-class correlation (ICC) was the statistic used to assess IRR. Unlike Cohen's kappa, which quantifies IRR based on all-or-nothing agreement, ICCs consider the magnitude of disagreement, with larger disagreements resulting in lower ICC values [20].

The selection of ICC parameters was guided by the study's design and the specific type of agreement we aimed to capture. In this study, the following ICC variants were chosen: i) two-way model, since the same raters evaluated all subjects, making it suitable for assessing both consistency and absolute agreement across raters; ii) mixed effects model, as coders were not randomly selected, the analysis focused on assessing the degree of agreement within the specific context of this study, rather than generalizing the results to a broader population of coders; iii) consistency type, because our main objective was to evaluate the consistency of coder ratings, focusing on how higher ratings by one coder corresponded with higher ratings by another, rather than requiring exact agreement in absolute values; iv) single measures unit, for the reason that the focus was on the reliability of individual ratings rather than the average across raters.

Moreover, since multiple variables were rated for each preservice teacher, separate ICC analyses were conducted for each variable. The interpretation of ICC values followed [21] commonly cited cutoffs for qualitative ratings of agreement: poor (<.40), fair (.40-.59), good (.60-.74), and excellent (.75-1.0).

3 RESULTS

3.1 Selection and Implementation of Data Collection Instruments to Map PCK Development

The selection and application of data collection instruments was a crucial process that demanded careful consideration, particularly with regard to the PCK components to be assessed, the timing of assessments throughout the master's program, and the choice of instruments. It's important to emphasize that these factors were not addressed in isolation; while the decisions will be presented in sequence, the actual process was highly iterative, with each decision shaping and refining the others in a dynamic, interconnected manner.

The literature review identified the key dimensions of PCK as curricular knowledge, knowledge about students, knowledge about teaching strategies, and knowledge about assessment. These critical components, widely recognized in PCK research, formed the foundation for evaluating PCK in this study.

Regarding the assessment timeline, although the goal was to map PCK development throughout the master program, we had to address a key concern: respondent burden. Response fatigue, resulting from the cognitive load and time required to complete the measurement tasks, can lead to lower response rates and incomplete data, ultimately reducing the richness and quality of the data. To mitigate this, we opted to collect data at three key points: at the beginning of the 2021/2022 academic year, corresponding to the start of the professional master's program; at the beginning of the 2022/2023 academic year, corresponding to the start of the second year of the same master's program; and after the end of the 2022/2023 academic year, following the participants' graduation.

For the selection of instruments, various factors were considered. As [22] highlight, individual interviews, while widely used for PCK data collection, are time-consuming and costly for both participants and researchers. Although we recognize the importance of interviews, we chose to conduct them only at the end of the program, aiming to identify the factors that participants believed had influenced their PCK development.

Recent findings [22] suggest that lesson plans (LP), being integral to teachers' daily practice, offer an accessible and cost-effective means of collecting PCK data from large samples, which aligns with our objectives. However, it also became evident that relying solely on LPs would be insufficient to capture the full depth and complexity of PCK. Therefore, in response to these conclusions, we decided to use two complementary instruments: CoRe (Content Representation) and LP.

The CoRe (Content Representation) tool [6] requires future teachers to select key ideas they believe are essential for students to learn a specific curricular topic. For each idea, participants explain its importance, provide the rationale for their pedagogical choices, consider students' understanding, and outline methods for assessing learning outcomes.

Thus, and following the approach of [7], we asked participants to design a 90-minute lesson plan for a 3rd-grade class on the topic of the respiratory system. This topic was chosen based on three criteria: its familiarity to students, its abstract nature, and its inclusion in the Portuguese curriculum. They were also asked to complete a CoRe on the same subject. Together, these instruments provided a comprehensive view of participants' PCK development.

3.2 Rubric analysis

The development of rubrics for measuring PCK is a relatively recent advancement, even so several options are already available. While using established research tools is generally considered best practice when assessing a well-researched topic like PCK, much of the existing literature presents tools that emphasize different aspects of PCK, often interpreted in diverse ways. Furthermore, it is essential to recognize that the analysis of a rubric cannot be separated from the examination of the instruments used to collect the data on which the rubric will be applied. So, based on previous decisions, rubrics focused on enacted PCK were excluded from consideration, as the project concentrated solely on PCK knowledge. In this context, two rubrics were analysed in greater detail, along with the instruments used to gather relevant data, namely the one reported by [23] and other by [24]. Both rubrics failed to accomplish our criteria. The former accomplish to measure only two PCK components and the latter rely heavily on individual interviews.

3.3 Rubric development: Descriptors

Based on the analysis of the rubrics conducted in the previous step, it seemed more appropriate to develop a rubric specifically tailored to the needs of our project.

For each dimension present in each of the four PCK components, specific performance criteria were established to capture the full spectrum of PST capabilities. Considerable effort was made to meticulously detail each aspect across all levels of the rubric.

The initial version of the rubric was subsequently reviewed by a panel of education experts, the authors, whose feedback was crucial in refining the criteria and descriptors. The second version (Table 1) was then employed to calculate the inter-rater reliability among coders. IRR was assessed using a two-way mixed, consistency, average-measures ICC to assess the degree that coders provided consistency in their ratings of empathy across subjects (Table 2).

The ICC values for the curriculum component are moderate, with goals (0.400) and curricular sequence (0.381) indicating some agreement but suggesting room for improvement in clarity and specificity.

In the component knowledge about students, there is a wide range of agreement. Alternative conceptions (0.692) and difficulties and common errors of students (0.770) show strong agreement, indicating that these areas are more clearly defined. However, two dimensions present particular challenges: prerequisites and variations in approaches to learning. For prerequisites, all coders assigned preservice teachers a level 1 rating. While one might expect lower levels of the rubric to have higher frequencies, this result requires more detailed analysis. It is possible that the way the instruments were applied does not allow for adequate data collection on this aspect, or it could be related to the fact that we are dealing with future teachers, indicating that this dimension may be underdeveloped in initial teacher training.

Table 1. Second version of the rubric.

Component	*Dimension*	*Level 1*	*Level 2*	*Level 3*	*Level 4*
Curriculum knowledge	Goals	The teacher does not present objectives, or the objectives are disconnected from the topic/make no sense	The teacher presents at least 1 appropriate objective (e.g. appropriate to the topic; age)	The teacher presents at least 2 appropriate objectives (e.g. appropriate to the topic; age)	The teacher presents at least 2 appropriate objectives, and the justification shows relevant knowledge (e.g. appropriate to the topic; age)
	Curricular sequence	The teacher does not make explicit any curricular knowledge	The teacher justifies the objectives in a too general way/mobilizes curricular knowledge implicitly (e.g. does not address very complex concepts)	The teacher justifies the objectives by making explicit 1 previous or subsequent learning predicted in the curriculum	The teacher justifies the objectives by making explicit more than 1 previous or subsequent learning predicted in the curriculum
Knowledge about students	Prerequisites	There is no reference to necessary knowledge or skills for learning	There is reference to 1 or more pieces of knowledge or skills necessary for learning, but they are generic	There is reference to 1 or more pieces of knowledge or skills necessary for learning a specific concept of the topic	There is reference to 1 or more pieces of knowledge or skills necessary for learning a specific concept of the topic, and it explains why it is important
	Alternative conceptions	There is no reference to alternative conceptions	There is reference to the existence of alternative conceptions, but none are made explicit	There is an explanation/description of 1 or 2 alternative conceptions, but they are not mobilized	There is an explanation/description of 1 or 2 alternative conceptions, and they are mobilized
	Difficulties and common errors of students	Does not indicate students' common difficulties/errors	Indicates 1 or 2 students' common difficulties/errors, but they are generic	Indicates 1 or 2 students' common difficulties/errors, specific to the topic	Indicates 1 or 2 students' common difficulties/errors, specific to the topic, and there is evidence that the teacher understands the origin of the difficulties/errors
	Variations in approaches to learning	There is no evidence that the teacher is aware that students can benefit from being exposed to different approaches to the same concepts	There is 1 reference that students can benefit from being exposed to different approaches to the same concepts	There are 2 references that students can benefit from being exposed to different approaches to the same concepts	There is at least 1 reference that students can benefit from being exposed to different approaches to the same concepts, and there is evidence that this knowledge is mobilized for teaching
Knowledge about teaching strategies	Specific strategies/ activities of the topic	The teacher does not refer to any specific science teaching strategy	The teacher refers to 1 specific science teaching strategy (experiments/scientific practices; use of models; concept maps; ... cognitive conflict)	The teacher refers to 2 specific science teaching strategies (experiments/scientific practices; use of models; concept maps; ... cognitive conflict)	The teacher refers to 3 or more specific science teaching strategies (experiments/scientific practices; use of models; concept maps; ... cognitive conflict)

	Appropriateness of the strategies/ activities mobilized	The teacher mobilizes only inappropriate strategies for the topic and/or in such a way that does not promote students' learning	The teacher mobilizes at least 1 activity/strategy that is not appropriate to the topic (so there are appropriate and inappropriate ones, regardless of the justification)	The teacher mobilizes only activities that are appropriate to the topic, but does not justify	The teacher mobilizes only activities that are appropriate to the topic and justifies based on how much they represent or clarify important information about the topic or associated scientific concepts
	Representations	The teacher does not refer to representations or refers to them incorrectly	The teacher refers to 1 representation correctly, but without justifying its advantages and disadvantages in supporting student learning	The teacher refers to 1 representation correctly and justifies its advantages and disadvantages in supporting student learning	The teacher refers to more than 1 representation correctly and justifies its advantages and disadvantages in supporting student learning
Knowledge about assessment	Assessment strategies	There are no references to assessment strategies	There is a reference to 1 type of assessment strategy	There is a reference to 2 or more types of assessment strategies	There is a reference to 2 or more types of assessment strategies that allow for student self-assessment
	Purpose of assessment (when and why)	The teacher does not assess	The teacher assesses only at one moment	The teacher assesses at more than 1 moment, but with only 1 purpose (diagnostic, formative, summative)	The teacher assesses throughout the class and with more than 1 purpose

. Table 2. IRR was assessed using a two-way mixed, consistency, average-measures ICC.

Component	***Dimension***	***ICC***
Curriculum saliency	Goals	0,400
	Curricular sequence	0,381
Knowledge about students	Prerequisites	0
	Alternative conceptions	0,692
	Difficulties and common errors of students	0,770
	Variations in approaches to learning	0,326
Knowledge about teaching strategies	Specific strategies/activities of the topic	0,817
	Appropriateness of the strategies/activities mobilized	0,706
	Representations	-0,333
Knowledge about assessment	Assessment strategies	-0,103
	Purpose of assessment (when and why)	-0,067

High ICC values in the knowledge about teaching strategies component reflect robust agreement among coders for specific strategies/activities (0.817) and appropriateness of strategies/activities (0.706), signifying a strong understanding of these teaching elements. In contrast, the ICC value for the representations dimension is notably low, suggesting significant discrepancies among coders. This may indicate issues with the clarity of the rubric for this dimension or difficulties in interpreting and evaluating this aspect consistently, highlighting the need for further refinement and training.

Conversely, both dimensions of knowledge about assessment — assessment strategies (-0.103) and purpose of assessment (-0.067) — exhibit significant disagreement, highlighting inconsistencies in coder interpretations and the need for clearer guidelines.

In summary, while some dimensions of the four components demonstrate strong reliability and agreement among coders (e.g., alternative conceptions, difficulties and common errors, and specific

strategies), others reveal significant inconsistencies that highlight areas for improvement in the rubric's clarity and effectiveness.

4 CONCLUSIONS

Pedagogical Content Knowledge (PCK) is a critical component of teachers' expertise, and its development is essential for enhancing teaching quality. However, PCK remains challenging to assess. A common approach to measure PCK is through the creation of scoring rubrics. This paper outlines the process of developing a scoring rubric tailored to evaluate PCK in a Portuguese master's program designed to prepare preservice elementary teachers.

One of the primary aspects to emphasize is the diversity of existing instruments and rubrics. There are instances where similar instruments are assigned different names or where the same term is used for distinct procedures. This inconsistency is concerning, as it may hinder the development of knowledge in the area of PCK.

The rubric development process involved several stages: instrument selection and data collection, the creation of performance criteria for each PCK component (through iterative discussions and literature review performed by the authors), and the calculation of inter-rater reliability among coders. While some PCK components, such as knowledge about students and knowledge about strategies, demonstrated fair to good inter-rater reliability, others, like curriculum saliency and knowledge about assessment, showed poor reliability. Additionally, certain dimensions, such as representations, exhibited very weak agreement.

In relation to the rubric, it is important to note that low inter-rater reliability in some dimensions of PCK can be attributed to a limited range, inadequately trained coders, or challenges in observing and quantifying PCK. To address these issues, a new discussion session will be organized with all coders to deliberate on whether to retain or eliminate dimensions and/or PCK components with low inter-rater reliability. If decision is made to keep at least some of them, then we should discuss reasons for disagreements and refine protocols. Additionally, a thorough review and retesting of the rubric is planned.

Ultimately, we believe that the output of the rubric, in conjunction with the process itself, will be highly beneficial for teacher training. The development of the rubric has proven to be a significant framework for discussing concepts and perceptions within a broader team involved in teacher education. Defining the quality of PCK knowledge and exploring how it translates into planning is an enriching and relevant endeavour. Furthermore, mapping PCK throughout initial teacher education can serve as a valuable tool for making evidence-based decisions. For example, it can help identify the most effective strategies for promoting the development of PCK during this critical period. Additionally, utilizing this instrument offers an overview of the class, enabling the trainer to identify areas for collective improvement while also addressing the individual needs of each student.

ACKNOWLEDGEMENTS

This work was funded by the Polytechnic Institute of Lisbon under the IDI&CA Contest (reference IPL/IDI&CA2023/PCK/PST_ESELx).

REFERENCES

[1] L. S. Shulman, “Those Who Understand: Knowledge Growth in Teaching,” *Educational Researcher*, vol. 15, no. 2, pp. 4–14, Feb. 1986, doi: 10.3102/0013189X015002004.

[2] M. Kunter, U. Klusmann, J. Baumert, D. Richter, T. Voss, and A. Hachfeld, “Professional competence of teachers: Effects on instructional quality and student development,” *J Educ Psychol*, vol. 105, no. 3, pp. 805–820, Aug. 2013, doi: 10.1037/a0032583.

[3] L. S. Shulman, “Knowledge and Teaching: Foundations of the New Reform,” *Harv Educ Rev*, vol. 57, no. 1, pp. 1–21, 1987.

[4] A. Berry, P. Friedrichsen, and J. Loughran, *Re-Examining Pedagogical Content Knowledge in Science Education*. Routledge, 2015.

[5] V. Kind, “Pedagogical content knowledge in science education: Perspectives and potential for progress,” *Stud Sci Educ*, vol. 45, no. 2, pp. 169–204, 2009, doi: 10.1080/03057260903142285.

[6] J. Loughran, P. Mulhall, and A. Berry, “In Search of Pedagogical Content Knowledge in Science: Developing Ways of Articulating and Documenting Professional Practice,” *J Res Sci Teach*, vol. 41, no. 4, pp. 370–391, Apr. 2004, doi: 10.1002/tea.20007.

[7] T. A. E. Van Der Valk and H. H. G. B. Broekman, “The lesson preparation method: A way of investigating pre-service teachers’ pedagogical content knowledge,” *European Journal of Teacher Education*, vol. 22, no. 1, pp. 11–22, 1999, doi: 10.1080/0261976990220102.

[8] A. Berry, F. Depaepe, and J. van Driel, “Pedagogical Content Knowledge in Teacher Education,” in *International Handbook of Teacher Education*, Singapore: Springer Singapore, 2016, pp. 347–386. doi: 10.1007/978-981-10-0366-0_9.

[9] C. Fernandez, “Knowledge Base for Teaching and Pedagogical Content Knowledge (PCK): Some Useful Models and Implications for Teachers’ Training,” *Problems of Education in the 21st Century*, vol. 60, no. 1, pp. 79–100, Jul. 2014, doi: 10.33225/pec/14.60.79.

[10] S. Magnusson, J. Krajcik, and H. Borko, “Nature, Sources, and Development of Pedagogical Content Knowledge for Science Teaching,” in *Examining Pedagogical Content Knowledge*, J. Gess-Newsome and N. G. Lederman, Eds., in Science & Technology Education Library. , Dordrecht: Kluwer Academic Publishers, 1999, pp. 95–132. doi: 10.1007/0-306-47217-1_4.

[11] S. Park and J. S. Oliver, “Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals,” *Res Sci Educ*, vol. 38, no. 3, pp. 261–284, 2008, doi: 10.1007/s11165-007-9049-6.

[12] S. Park, J. Suh, and K. Seo, “Development and Validation of Measures of Secondary Science Teachers’ PCK for Teaching Photosynthesis,” *Res Sci Educ*, vol. 48, no. 3, pp. 549–573, Jun. 2018, doi: 10.1007/s11165-016-9578-y.

[13] S. K. Abell, “Twenty Years Later: Does pedagogical content knowledge remain a useful idea?,” *Int J Sci Educ*, vol. 30, no. 10, pp. 1405–1416, Aug. 2008, doi: 10.1080/09500690802187041.

[14] A. Hume, R. Cooper, and A. Borowski, *Repositioning pedagogical content knowledge in teachers’ knowledge for teaching science*. Springer Singapore, 2019. doi: 10.1007/978-981-13-5898-2.

[15] J. Barnett and D. Hodson, “Pedagogical context knowledge: Toward a fuller understanding of what good science teachers know,” *Sci Educ*, vol. 85, no. 4, pp. 426–453, Jul. 2001, doi: 10.1002/sce.1017.

[16] K. F. Cochran, J. A. DeRuiter, and R. A. King, “Pedagogical Content Knowing: An Integrative Model for Teacher Preparation,” *J Teach Educ*, vol. 44, no. 4, pp. 263–272, Sep. 1993, doi: 10.1177/0022487193044004004.

[17] K. Carter, “Teachers’ knowledge and learning to teach,” in *Handbook of research on teacher education*, W. R. Houston and M. H. J. Sikula, Eds., NewYork: Macmillan, 1990, pp. 291–310.

[18] J. H. Van Driel, N. Verloop, and W. De Vos, “Developing Science Teachers’ Pedagogical Content Knowledge,” 1998.

[19] K. K. H. Chan and A. Hume, “Towards a Consensus Model: Literature Review of How Science Teachers’ Pedagogical Content Knowledge Is Investigated in Empirical Studies,” in *Repositioning Pedagogical Content Knowledge in Teachers’ Knowledge for Teaching Science*, Singapore: Springer Singapore, 2019, pp. 3–76. doi: 10.1007/978-981-13-5898-2_1.

[20] K. A. Hallgren, “Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial,” *Tutor Quant Methods Psychol*, vol. 8, no. 1, pp. 23–34, Feb. 2012, doi: 10.20982/tqmp.08.1.p023.

[21] D. V. Cicchetti, “Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology.,” *Psychol Assess*, vol. 6, no. 4, pp. 284–290, Dec. 1994, doi: 10.1037/1040-3590.6.4.284.

[22] Z. Kaya and O. N. Kaya, “Gathering Rich Data on Preservice Science Teachers’ Pedagogical Content Knowledge Through Their Lesson Plans,” *J Teach Educ*, vol. 74, no. 1, pp. 10–22, Jan. 2023, doi: 10.1177/00224871221105801.

[23] K. K. H. Chan, M. Rollnick, and J. Gess-Newsome, “A Grand Rubric for Measuring Science Teachers’ Pedagogical Content Knowledge,” in *Repositioning Pedagogical Content Knowledge in Teachers’ Knowledge for Teaching Science*, A. Hume, R. Cooper, and A. Borowski, Eds., Singapore: Springer Nature Singapore, 2019, pp. 253–271. doi: 10.1007/978-981-13-5898-2_11.

[24] D. L. Hanuscin, D. Cisterna, and K. Lipsitz, “Elementary Teachers’ Pedagogical Content Knowledge for Teaching Structure and Properties of Matter,” *J Sci Teacher Educ*, vol. 29, no. 8, pp. 665–692, Nov. 2018, doi: 10.1080/1046560X.2018.1488486.