Research in Science Education
https://doi.org/10.1007/s11165-022-10042-0

Developing an Instrument to Assess Pedagogical Content Knowledge for Evolution

Beatriz Becerra[1] · Paola Núñez[1] · Claudia Vergara[2] · David Santibáñez[3] · Dirk Krüger[4] · Hernán Cofré[1,5]

Accepted: 5 January 2022

Abstract
Despite the importance of evolution to understand biology, there is significant evidence that many biology teachers have difficulties to successfully teach this topic. The purpose of this study is to describe procedures by which a paper-and-pencil instrument to assess teachers' pedagogical content knowledge for evolution (PCK_{evo}) was developed and validated. The instrument was created to measure the components proposed in the model of Magnusson et al. (1999), except for the orientations. The creation of the questionnaire was carried out through eleven steps: (1) conceptualization, (2–4) creation of items and rubric reviewed by experts (focus on content and construct validity), (5–7) piloting a 32-item questionnaire to 10 biology teachers (focus on face and construct validity and reliability), (8–10) piloting a 16-item questionnaire to 61 biology teachers (focus on discriminant and construct validity and reliability), (11) application of a different instrument (CoRe interview) to four teachers (focus on criterion validity). The final version of the questionnaire (which includes three PCK components) showed in a Rasch analysis for the reliability of the items ($\alpha = 0.90$) and persons ($\alpha = 0.81$) adequate values. When applying the final version of the questionnaire, there is also evidence of discriminant validity (differences between two groups of teachers with or without professional development in evolution education). The contributions of the PCK_{evo} instrument for research topics and evolution teaching are discussed.

Keywords Evolution · Natural selection · Paper-and-pencil instrument · Pedagogical content knowledge

Introduction

There is a significant amount of literature that confirms that evolution and its main mechanism natural selection is poorly understood at all levels, including elementary students and high school students (e.g., Kampourakis & Zogza, 2007; Cofré et al. 2018a), biology majors (e.g., Cofré et al. 2016), pre-service biology teachers (e.g., Cofré et al. 2016;

✉ Hernán Cofré
hernan.cofre@pucv.cl

Extended author information available on the last page of the article

Friedrichsen et al., 2018), and even in-service biology teachers (e.g., Nehm & Schonfeld, 2007; Cofré, Cuevas & Becerra 2017).

One explanation offered for students' low acceptance and low understanding of evolution is the weaknesses of biology teachers in teaching the content. There is significant evidence that describes the difficulties biology teachers have regarding their beliefs, limitations, and insecurities about evolution (Sickel & Friedrichsen, 2013; Glaze & Goldston, 2015; Deniz & Borgerding, 2018). Because of the key role of teachers in the creation of significant learning on evolution, the question arises regarding how prepared biology teachers are to teach this important content (e.g., van Dijk 2009; Cofré et al. 2017; Deniz & Borgerding, 2018). More specifically, it is important to study which is the pedagogical content knowledge (PCK) of evolution that biology teachers display, this is, what knowledge do they have regarding how to teach and evaluate this subject, and which are the students' learning difficulties (e.g., Cofré et al. 2016). One of the limitations that prevents teachers from adequately teaching this content is the lack of training on teaching strategies to teach evolution (e.g., Romine et al., 2014; Sickel & Friedrichsen, 2013; Cofré et al. 2017) and to work with student preconceptions (Lucero et al., 2017; Parraguez et al. 2022). Despite taking this under consideration, there are some gaps in the research related to the knowledge about teaching and understanding evolution, specifically on how to evaluate it and how it relates to students' knowledge and acceptance of evolution (Nehm & Schonfeld, 2007; Sickel & Friedrichsen, 2017; Lucero et al., 2017).

On the other hand, PCK has been considered for several decades to be one of the most important components of professional knowledge, but also one of the most difficult to understand (van Driel et al., 2014; Gess-Newsome, 2015; Chan & Hume, 2019). In recent years, the great development in research about PCK of science teachers has centered around several subjects, one of those is how to measure or evaluate it (e.g., Alonzo & Kim, 2016; Chan & Hume, 2019). However, the studies that describe reliable and valid instruments to evaluate PCK in biology teachers are still few, especially in evolution (Chan & Hume, 2019; Großschedl et al., 2019). In this context, the main goal of this study is to describe procedures by which a paper-and-pencil instrument to assess teachers' PCK for evolution (PCK_{evo}) was developed and validated.

Literature Review

Conceptualization and Measuring of Pedagogical Content Knowledge

Shulman (1986) proposed his ideas about how the topics of any scientific content should be addressed in class, and how the teacher should be aware of different elements that would be the key to generate learning in students. Magnusson, Krajcik, & Borko (1999) state that knowledge pertaining to teaching is composed by five components that can be measured separately: (1) orientations towards science development, (2) knowledge and beliefs about the science curriculum (KSC), (3) knowledge and beliefs about student understanding of science specific issues (KSU), (4) knowledge of assessment of science learning (KAS), (5) knowledge and beliefs about instructional strategies and representations (KISR). During the last decade, a search began to find a model able to unify many ideas previously worked in parallel by different researchers. This is how Gess-Newsome (2015) proposes a consensus model (CM) thought from a professional development perspective, which combines content, pedagogy, and context. Despite the contributions of this model, the search continued for a new, more robust model in theoretical and practical terms, suggesting in

2019 the Refined Consensus Model (RCM; Carlson & Daehler, 2019). Despite the current existence of these two models, the model of Magnusson et al. (1999) continues to be the most universally used in the literature according to some of the most cited reviews on PCK (e.g., Abell, 2007; van Driel, Berry, & Meirink, 2014; Chan & Hume, 2019). In fact, Chan & Hume, 2019 have showed that from 100 articles published on PCK in the main science education journals between 2008 and 2018, about 50% have used Magnusson's model or modifications of it. One of the most widely used variations of this model is the proposal by Park & Oliver (2008) who has focused on measuring the interactions of the five original components of Magnusson's PCK model (e.g., Park & Suh, 2015). Thanks to this last model, it has been described that the different knowledge within the PCK can interact, develop differently, and depend on the content (Park & Chen, 2012; Bravo & Cofré 2016; Reynolds & Park, 2021). Due to the CM and RCM eliminate these components and thus the possibility of continuing to better understand the development and interaction of this knowledge, this research will use Magnusson's PCK model proposal as the theoretical framework.

On the other hand, PCK measuring requires complex and special procedures given that it involves assessing tacit procedural knowledge. For this reason, different reviews have proposed that the best way to measure and describe PCK is to collect data from different sources (e.g., Baxter & Lederman 1999; Chan & Hume, 2019). In their review, Chan & Hume (2019) determined that research done in the last 10 years frequently had used more than one sources of data to gather information about PCK from science teachers. These approaches often resort to triangulation of structured and semi-structured interviews, stimulated recall interviews, artifacts from teachers task, concept mapping, teachers reflective journals, and lessons observations (e.g., Park et al., 2011; Jüttner & Neuhaus, 2012; Rollnick, 2017; Coetzee, Rollnick & Gaigher, 2020; Reynolds & Park, 2021). One of the most widely used proposals to measure PCK has been the CoRes (Content Representation) and PaP-eRs (Pedagogical and Professional-experience repertories) procedures developed by John Loughran and his colleagues (e.g., Loughran, Mulhall & Berry, 2004, 2008). These instruments have been widely used to describe and document the PCK of secondary and elementary teachers (e.g., Loughran, et al., 2004, 2008; Padilla et al., 2008; Bravo & Cofré 2016; Rollnick, 2017; Coetzee, Rollnick & Gaigher, 2020). The main outcome of these efforts has been the conviction that the best way to characterize PCK is both through observation of teacher performance and interviews where the teacher explains why he or she made the pedagogical decisions observed. However, despite the good results of these approaches, qualitative studies tend to include small samples and require significant time for its analysis. On the other hand, although the use of questionnaires and tests has been one of the most used methodologies in the study of PCK in recent years, no more than six articles have focus on develop questionnaires to determine PCK in science teachers (Chan & Hume, 2019). Paper and pencil evaluations have been described as a promising tool to determine PCK, since it is a reliable, objective, and valid method that can be applied to larger samples (Schmelzing et al., 2013) and can be also used in conjunction with teaching artifacts (e.g., Kanter & Konstantopoulos, 2010), interviews (e.g., Zhou et al., 2016), or lesson observations (e.g., Roth et al., 2011).

Pedagogical Content Knowledge on Evolution

Although evolution is a complex subject to teach, research about PCK_{evo} in teachers is scarce (e.g., Lucero et al., 2017; Cofré et al. 2016, 2017; Friedrichsen et al., 2018;

Sickel & Friedrichsen, 2018, see also Sickel & Friedrichsen, 2013 for a review about teacher knowledge about teaching evolution). Among these, the study of Lucero et al. (2017) included four teachers at a secondary school in the USA, which revealed that teachers were not making use of alternative conceptions of their students to direct their teaching, and these conceptions were only corrected or not acknowledged. In a different study, Bravo & Cofré (2016) explore how PCK_{evo} of two teachers is modified by participating in a professional development program. Through qualitative methods such as CoRes and PaP-eRs, the results show that reflective processes about the teaching practice promote transformation of beliefs and knowledges about optimal teaching methods and strategies of evolution. In another study, Sickel & Friedrichsen (2018) decided to conceptualize the PCK development of three biology teachers while they taught natural selection. Using three different analytical lenses to evaluate PCK_{evo}, researchers concluded that teacher knowledge and content preparation affect teaching. A common aspect of these studies has been the qualitative methodology that has been used to study PCK_{evo}, which results in small samples and prevents generalization. Therefore, performing quantitative studies is a step forward in the research of PCK_{evo}.

Understanding and Teaching Evolution

The obstacles that can be observed when teaching and understanding evolution are related to multiple issues: teleological thinking (species evolve out of necessity to adapt to the environment), alternative conceptions about how science works, media expressions and terminology, religious beliefs of students and the same mistakes that are present in biology classes as teachers address the content (Nehm, 2018). The teleological idea, for example, is a preconception that tends to repeat itself in a great number of students, where the thought is that biological change in species occurs by the will of individuals to adapt to the environment (Kampourakis & Zogza, 2007; Nehm, 2018). On the other hand, for a long time, preconceptions about the nature of science have made acceptance and understanding of the theory of evolution more difficult for students. In their review, Glaze & Goldston (2015) indicate that both acceptance and understanding of evolution have a positive relationship with the nature of science. However, other studies have proposed that this relationship could be more complex (Cofré et al. 2017). These authors showed that, in their study with 31 biology teachers that participated in a professional development program, a significant correlation was not found between the understanding of the natural selection mechanism, acceptance of evolution and the understanding of nature of science by teachers.

On the issue of teaching evolution, Glaze & Goldston (2015) in a thorough review of the topic during the last years state that student-centered teaching, which includes active learning would be the most effective approach. For example, the research carried out by Cofré et al. (2018a, b) demonstrates that doing activities that work with conceptual change and emphasize the use of real data from actual research about evolutionary phenomena significantly improved eleventh grade students' understanding of the mechanism of natural selection. Moreover, the use of models and laboratory activities focused on socio-scientific issues (e.g., bacteria resistance to antibiotics) allows to increase the understanding of the evolutionary mechanism of natural selection, and in addition, it would not only yield good results for improving understanding, but these types of strategies also allow the development of scientific thinking abilities such as data analysis and argumentation (Williams et al., 2018).

Research Questions and Hypothesis

In accordance with the reviewed literature, the following questions guided this study: (1) What validation and reliability processes should an instrument undergo to adequately describe PCK_{evo} in biology teachers? (2) What will be the discriminant power of the instrument if PCK_{evo} is evaluated in biology teachers with and without experience in teacher professional development on evolution? Therefore, the hypotheses are as follows: (H1) The inclusion of varied and pertinent sources of validation, such as those indicated in the literature, will allow the instrument to adequately approximate the PCK_{evo} of biology teachers; (H2) Teachers that have participated in a professional development program related to the teaching of evolution will score higher on the PCK than teachers that have not received improvement training in this area.

Methods

Construction of an Instrument to Approach PCK_{evo}

The detailed identification of the underlying theory of the research is one of the key points for the construction of a reliable instrument with solid evidence of the validation processes that can capture PCK_{evo} for biology teachers. Following sections present eleven steps in the construction of the questionnaire developed to measure PCK_{evo} (Fig. 1). The questionnaire was created using similar procedures to those described in the literature related to PCK measuring (e.g., Jüttner & Neuhaus, 2012; Schmelzing et al., 2013; Jin et al., 2015; Kirschner et al., 2016; Park et al., 2018; Großschedl et al., 2019).

Step 1. Conceptualization of PCK.

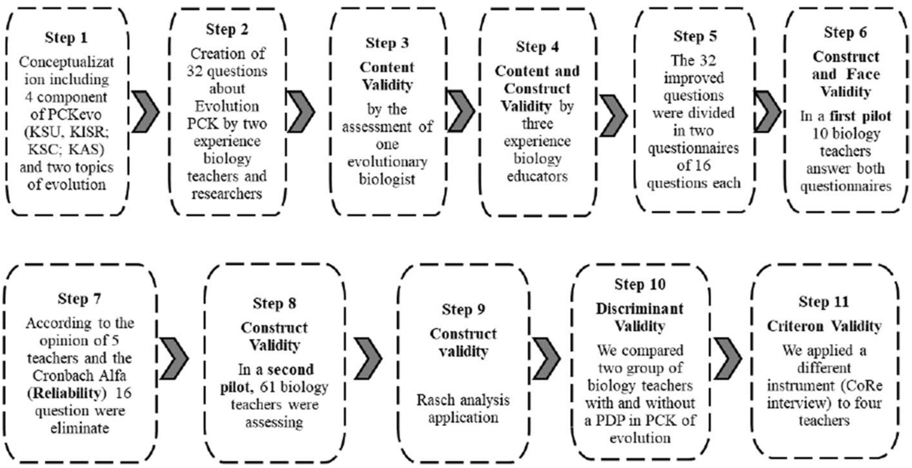


Fig. 1 Schematic diagram of the eleven steps to create the instrument on PCK of evolution

The instrument was developed including evolutionary topics, present at Chilean curriculum (Mineduc, 2015, 2019) as well as, at international literature (e.g., Deniz & Borgerding, 2018). In both, there is emphasis on the evidence supporting evolution to grant it valid scientific knowledge status and consider natural selection as a mechanism of the origin of species. Regarding natural selection, literature has pointed out that this mechanism presents significant difficulty for its teaching, for both students and teachers (e.g., Lucero et al., 2017; Nehm, 2018). On the other hand, the issue of evidence for evolution has also been recognized as an important point to address due to its implications in the acknowledgement of the scientific status of the theory of evolution (e.g., Glaze & Goldston, 2015; Cofré et al. 2018a, b). Regarding PCK conceptualization, questions in the questionnaire were created considering four components of the model of Magnusson et al., (1999): KSU, KISR, KSC, KAS. For each component, eight questions were created, which related to natural selection and theory and evidence evaluated by the instrument.

Step 2–4. Creation of Items and Rubric.

Considering the guidelines from Step 1, two biology teachers with experience in research and classroom teaching of evolution created eight open-ended questions for each PCK component (32 questions in total) (Table 1). The teachers resorted to literature to elaborate items related to teaching strategies for difficult topics and alternative conceptions of students about evolution (e.g., Robbins & Roy, 2007; Williams et al., 2018). The validity of the construct was examined by two researchers with broad experience in teaching these contents in schools, universities, and in-service teacher training. The validity of the content was reviewed by an evolutionary biologist of well-known trajectory both in research and university teaching of evolution. To categorize the answers from teachers, we elaborated an analytic rubric that classifies their answers in three possible categories and where the score is obtained according to the level in which the answer is classified (Table 2). The rubric was created deductively, including knowledge about PCK_{evo} from the literature, including for example the use of preconceptions for planning instruction or the use of good teaching strategies such as include nature of science or human evolution (e.g., Sickel & Friedrichsen, 2013; Cofré et al. 2016, 2017; Friedrichsen et al., 2018; Harms & Reiss, 2019). In level 2 (Table 2), answers are explicit, complete, and integrate all quality criteria mentioned in the question. In level 1 (Table 2), answers included one of the quality criteria in the highest category or those that included enough criteria but presented ambiguities that made it difficult to clearly

Table 1 Example of instrument item by component. The whole instrument is available in Supplementary material

Component	Items
KSU	When performing an assessment on evolution after addressing the content and asking students about their position on the evolution theory, you realize that many of them understand it, but do not accept it. Why do you think that, despite the evidence presented in class, there are students that still do not accept it? Explain
KISR	During evolution class, a student comments that certain plants had to develop thorns in their stem to avoid being eaten by herbivores. According to this information, what questions would you ask the student to confront him with this alternative conception?
KSC	Explain why it is important to know the essentials of DNA to address the concepts of Evolution and Biodiversity

Table 2 Example of analysis response using the created rubric. The words in bold correspond to the quality elements mentioned by teachers

Component: knowledge of strategies

Item	Problem statement	0 points	1 point	2 points
11	Silvia and Susana meet to plan evolution lessons. Silvia tells Susana that she considers convenient to talk to students, in addition to the evolution theory, about other thoughts on the origin of species such as creationism, transformism, and fixism (as stated in the national study plan). However, Susana thinks this lesson would not be convenient since it could lead to mistaken ideas in students. Which teacher do you agree with and why? Explain	I think Silvia is right because she approaches her students with her previous knowledge such as creationism	Personally, I think they should mention them. Also, since said theories have neither empirical support nor a supporting methodology beyond the account and characteristics they may have, as science teachers, we cannot go in-depth into these theories, leaving a rather unfinished debate	With Silvia, because addressing other proposals considers addressing aspects of nature of science, as well as its social and cultural construction, in addition to allowing to differentiate the evolution theory from other forms of non-scientific knowledge. With these considerations, evolution classes become a powerful opportunity to strengthen the discussion and development of critical thinking in students, an opportunity that is lost according to Susana's proposal

understand their decision, thought, or position. Finally, in level 0 (Table 2), answers did not include any criteria mentioned at the highest level or did not indicate any of comprehension of the problem or question. The rubric was elaborated by the same teachers and researchers that created and validated the instrument.

Step 5–7. Validity and Reliability of Pilot 1.

Ten biology teachers participated in pilot 1. Most teachers had additional training in evolution education, through workshops and courses of teacher development programs. Two questionnaires of 16 questions each (32 questions in total) were applied online to the teachers (one questionnaire at a time), with a 1-month deadline to complete both instruments. In both questionnaires, there was a section to comment on the question and contribute in this way to the improvement of the items of the instrument (construct validity). With the results of the analysis, the Cronbach's alpha was calculated as a measure of internal consistency for the complete questionnaire ($\alpha = 0.63$) as well as for each PCK component (KISR: $\alpha = 0.64$; KSU: $\alpha = 0.45$; KSC: $\alpha = 0.65$; KAS: $\alpha = 0.33$). Here, we are understanding reliability or internal consistency as the extent to which an instrument can be expected to give the same measured outcome when measurements are repeated (Taber, 2018). From the results of pilot 1 and considering the feedback by teachers, 16 questions were eliminated. This process excluded the PCK component knowledge of assessment of science learning (KAS) and reduced the number of questions per component.

Step 8–10

a) Validity and Reliability of Pilot 2.

Discriminant validity is the validity obtained from the instrument when comparing two groups that, according to their nature, are expected to obtain different results or what is the same, the instrument allows "discriminating" between said groups (AERA, APA, and NCME 2014; Ergönenç, Neumann & Fischer 2014; Lederman et al. al., 2014; Großschedl, et al., 2019). Pilot 2 included a group of 61 teachers with and without participation in professional development programs (PDP) on understanding and teaching evolution, which provided an opportunity to perform a comparison between their performances (discriminant validity). Based on existing evidence about the effectivity of the PDP in which they participated is expected a difference between both groups (Cofré et al. 2017, 2018a, b). Of the 61 teachers, 16 of them had participated in PDP which included general aspects of science education, knowledge, and teaching of nature of science, and knowledge and teaching of evolution (see Cofré et al. 2017 for more details on PDP). Teachers who participated in this study were in the 21 to 55-year-old range and had a professional experience of 1 to 25 years; 72% of them are female and 28% are male. Teachers had 2 weeks to respond the questionnaire at home. In this phase of the study, participant teachers were rewarded with three different types of presents: a book on the subject (evolution or genetics), hands-on resources for the teaching of evolution, or a flash drive with evolution teaching materials (scientific articles, activities, power point presentations, etc.). Two authors independently coded 976 responses using the rubric. Mean inter-rater reliability values of questions was measured using Kappa ($k = 0.76$). Three questions present a moderate Kappa value between ($k = 0.40$–0.60); five questions present a satisfactory Kappa value ($k = 0.61$–0.80); and eight questions present an extremely high Kappa value ($k = 0.81$–1). Furthermore, a Spearman's correlation analysis between the total score by teacher obtained by each researcher reached a value of $r = 0.93$ ($p < 0.001$). This shows that the discrepancies in the three lowest performing questions do not affect the final PCK value of each biology teacher. Discordant scores were resolved via deliberation between the two researchers, leading to a set of consensus scores for all concepts.

b) Data Analysis (Pilot 2).

For data analysis of pilot 2, we also performed the Cronbach's alpha test with the SPSS 20.0 program. Furthermore, and as a complementary statistical measure (Taber, 2018), it was used the partial credit model by Rasch (Andrich, 1978), an analysis reported by various studies related to the creation of instruments to measure PCK (e.g., Jüttner & Neuhaus, 2012; Kirschner et al., 2016; Großschedl et al., 2019). Rasch's measuring offers a broad range of techniques that can be used to evaluate the validity and reliability of evaluation instruments. A distinctive feature of this analysis is that it offers a joint measuring, this is, the parameters of people and items are expressed in the same units and are in the same continuum, which allows to visualize the interaction of the participants of the study with the questions of the instrument (Andrich, 1978). To prove the items' fit to the Rasch model, the sum of the squared standardized residuals (MNSQ) were used. The MNSQs, which indicate how accurately data of each item fit the respective item response theory's assumptions, have acceptable values when ranging from 0.5 to 1.5 (Boone et al., 2014). The validity of the test construct was evaluated using the Wright map (Aryadoust, 2009; Boone et al., 2014). For the instrument to be validated by the results obtained after the application of Rasch, the average of the results from teachers should be close to the average of the instrument's items, which would indicate a good orientation of the items. Likewise, the reliability of both should be close in relation to one another, and each one of them should be > 0.7 (Jüttner & Neuhaus, 2012).

Step 11. Criterion Validity.

Criterion validity, also known as concurrent validity and prognostic validity, compares the results of an instrument with a criterion assessed at the same time (Fischer et al., 2014). A qualitative method to capture PCK of science teachers that is widely used in the literature are CoRe and PaP-eR (Loughran et al., 2004; 2012 see Literature Review). Here, the CoRe was applied to four teachers that answered the PCK_{evo} in pilot 2, with the goal of evaluating the existence of correspondence between the teacher's score in the questionnaire and the description of the knowledge declared on the CoRe, this is, we evaluated PCK of four teachers with a different criterion at the same time (Fischer et al., 2014). To achieve a clear representation of what is presented in the CoRe, we decided to follow the methodology presented in the work of Bravo & Cofré (2016). Specifically, the CoRe diagrams generated for each teacher were described highlighting 4 traits: (a) the total number of concepts present in the diagram, (b) the number of connections between concepts and the diagram, (c) number of new concept or subcategories inside PCK components, and (d) the number of concepts of each teacher agree with the knowledge expressed in a collective CoRe created by a group of six expert biology teachers and biology educators (Cofré et al in preparation). As describes by Carlson & Daehler (2019), this could be a valid way of capture the collective PCK about evolution.

Results

Cronbach's Alpha Reliability Analysis

The results show that there was a change in the values of the Cronbach's alpha for the whole instrument increased from $\alpha = 0.63$ to $\alpha = 0.76$ in pilot 2, being this a good value for an internal consistency in PCK questionnaires (e.g., Ergönenç et al., 2014; Park et al., 2018). Among the evaluated PCK components, the Cronbach's alpha obtained were almost equal (KISR: $\alpha = 0.65$; KSU: $\alpha = 0.46$; KSC: $\alpha = 0.65$). Although, one of these values is low, the small number of items in each of the PCK subcomponents must be considered (Taber, 2018). On the other hand, the high Cronbach's alpha value for the entire instrument may reflect the internal consistency of the PCK as construct and the multiple interactions that exist between the PCK subcomponents (e.g., Park & Oliver, 2008) and not just an increase in the number of items (Taber, 2018).

Psychometric Analysis of the PCK Instrument

The average difficulty of the items was established at 0 for the adjustment test of the model, and therefore, the standard deviation comes close to the value 1, since most of the elements are between − 1 and + 1 logist ($SD = 0.5$). Regarding the infit and outfit MNSQ, the values of 15 items are within the expected interval ($0.5 < \text{MNSQ} < 1.5$); therefore, item (no. 16) was excluded of the analysis (Boone et al., 2014). For 15 items of the PCK evolution instrument, the reliability measure of the items ($rel._{EAP/PV} = 0.90$) and for persons ($rel._{EAP/PV} = 0.81$) indicate a productive measuring of the instrument (Großschedl et al., 2019). The Wright map of the PCK_{evo} questionnaire was used to analyze the validity of the construct of the instrument (Boone & Rogan, 2005). Figure 2 shows the Wright map of the PCK_{evo} questionnaire. Teachers (left side) that are at the bottom of the map are those with the lowest ability level, and in turn, those teachers at the top present a high ability to respond correctly to most questions on the instrument. On the bottom of the map (right side), there are those questions that were easy to answer and as we approach the superior section, questions that were hard to answer. The map indicates that there is an optimized measuring precision in the construction of the instrument. The final version of the PCK_{evo} questionnaire is included in the Supplementary material 1.

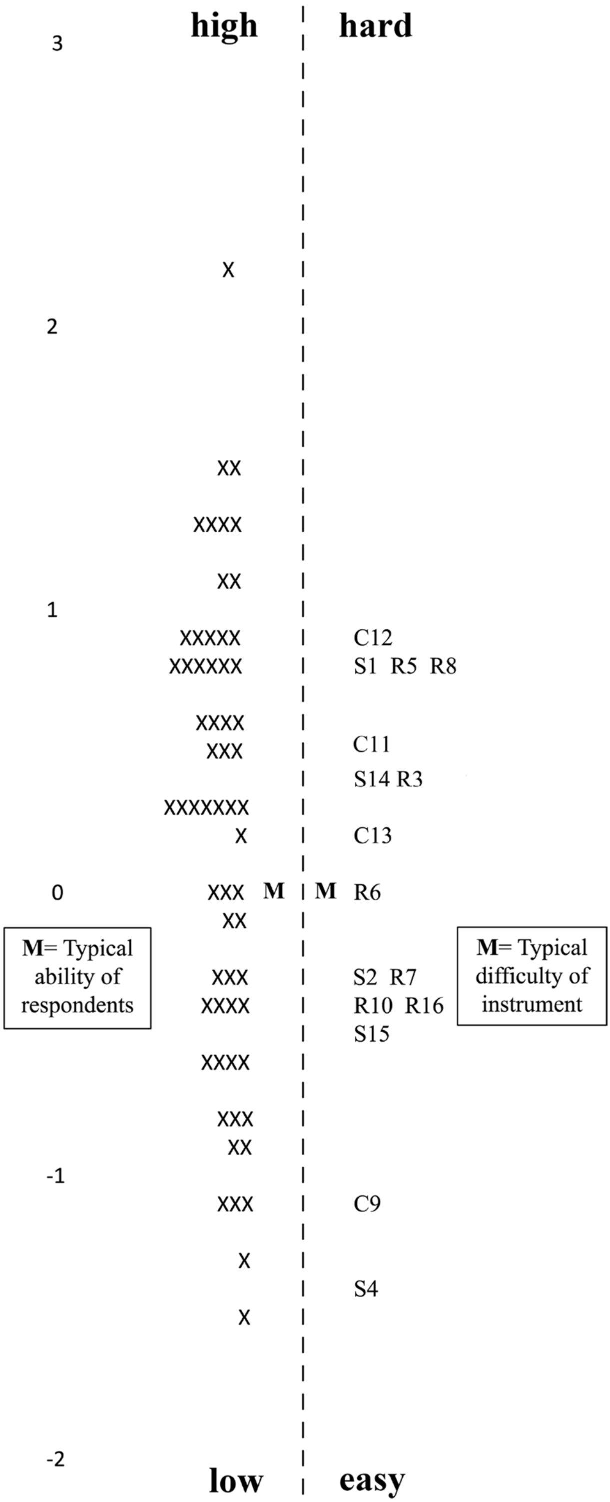
Comparison of Results Between 2 Groups (With and Without PDP)

Figure 3 reveals a significant difference between the scores obtained by teachers with and without experience in a professional development program about evolution and its teaching. The Mann–Whitney test between the two groups indicated that biology teachers that participated in the PDP have, on average, an PCK_{evo} value significantly higher than teachers without this experience ($z = -4.23$; $p < 0.001$; $r = 0.58$). Teachers that were part of the group that had not gone through evolution, PDP reached an average of 15.6 points, while those that had been a participating in PDP reached an average score of 21.9 points (Fig. 3).

Application of the CoRe Interview as an Evidence of Criterion Validity

A valid and reliable instrument should yield results that agree with previously recognized methodologies. In this case, when we compare the CoRe of four biology described a similar pattern: teachers with a diagram describing a more developed PCK obtained a questionnaire score considerably higher than that of the teacher with less complex diagrams. Specifically, diagrams belong to teacher with higher questionnaire score have more concepts, more connections among concepts, more concepts shared with the collective PCK for evolution, and present more exclusive subcategories. Additional concepts and subcategories are related to topics that literature on evolution PCK has acknowledged as important: *The*

Fig. 2 Wright Map for the results of PCK_{evo} ($n = 61$). Results of people are marked with an "X," those of questions are labeled with the question number and the type of knowledge evaluated was assigned a letter (R = KISR; S = KSU; C = KSC). Questions on the top part of the map represent the most difficult questions. The questions at the bottom of the map represent those easier to answer. Teachers that appear at the top part of the map are those who present a greater ability to respond to the items and those at the bottom of the map are those that have less ability to respond to the items



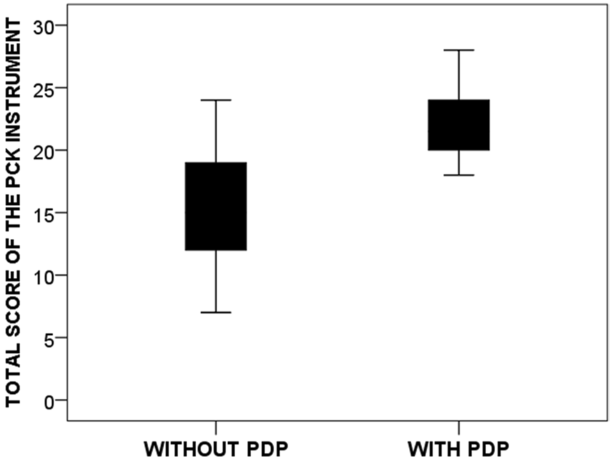


Fig. 3 Comparison between total mean results of teacher responses with and without participating in an evolving professional development program (PDP)

link with nature of science (NOS) (specifically how to teach the difference between theory and law to students) and teaching evolution *theory after reviewing essential genetics and ecology elements.* Supplementary material 2 shows the diagram of two biology teachers.

Discussion

Validation and Reliability Processes in the Preparation of an Instrument

During the last 7 years, only five articles in the field of teaching biology are related to the creation of instruments that will allow to capture PCK of a given biological content (Jüttner & Neuhaus, 2012; Jin et al., 2015; Park et al., 2018; Schmelzing et al., 2013; Großschedl et al., 2019), but none of them related specifically to evolution. Only Großschedl et al. (2019) included evolution as one of the six contents included in their biology PCK instrument. Therefore, the creation of a questionnaire of this kind represents a contribution to both PCK research, and research that seeks to improve the understanding and teaching of evolution. Regarding the information of the components the instrument provides, it evaluates three components of PCK, knowledge and beliefs about the science curriculum (KSC), knowledge and beliefs about student understanding of science specific issues (KSU), and knowledge and beliefs about instructional strategies and representations (KISR). KSU is not a feature usually included in instruments, since most of them evaluate only KSK and KISR (e.g., Jüttner & Neuhaus, 2012; Jin et al., 2015; Park et al., 2018; Schmelzing et al., 2013); however, we consider that the evaluation of KSU to be very important when you want to teach complex content such as evolution. Only Großschedl's et al. (2019) instrument has evaluated more components, which shows the difficulty in achieving this objective (Großschedl et al., 2019).

Regarding the hypotheses, the use of varied and pertinent sources contributed to the creation of an instrument, which, according to the results obtained in the validation processes, allows us to adequately approach the PCK_{evo} of biology teachers; therefore, the first hypothesis is accepted. The discriminant validity investigated with teachers with and without PDP on evolution meets the standards in international research for PCK instruments (e.g., Schmelzing et al., 2013; Kirschner et al., 2016; Großschedl et al., 2019). The results obtained by the comparison between the two interest groups in this study strengthen the results of the PCK instrument due to the discriminating power it had on the knowledge of the teaching evolution. Those teachers who had participated in a PDP on evolution obtained a significantly higher score than those without PDP, thus the second study hypothesis is also accepted.

The results of the Cronbach's alpha reliability analysis for the whole instrument show a good value within the literature about PCK measuring (e.g., Ergönenç et al., 2014). For example, Park et al. (2018) report a Cronbach's alpha like the one obtained in this study ($\alpha = 0.84$), while Großschedl et al. (2019) report a slightly higher value ($\alpha = 0.92$) and Jüttner & Neuhaus (2012) obtained a lower Cronbach's alpha value ($0.60 < \alpha < 0.65$). The use of the Rasch analysis was useful to connect theory with measuring helping in the selection of elements that would improve to measure PCK_{evo}. The interpretation of the results provided by the Rasch analysis supports the idea that the 15 item questionnaire offers a scoring with adequate reliability and a sufficient degree of validity. The Wright map, for its part, suggests that the ability of the teacher to respond to questions is distributed in a partially normal way, with a slight tail towards the respondents with the lowest ability level. Furthermore, the Wright map indicates that the items or questions are distributed close to typical question difficulty. This is relevant information for those researchers that wish to approach PCK_{evo} using an instrument with adequate psychometric properties (Großschedl et al., 2019). Großschedl et al. (2019) state that cross validation provides additional information about an instrument, and therefore, during the process of creation of this questionnaire, various resources were used which collaborated in the consolidation of the final instrument.

Recommendations and Implications for Future Research

The performed and reviewed evaluations suggest this questionnaire provides with an objective scoring, evidence for each validation process, and appropriate reliability. However, one of the limitations is related to the unequal number of questions per PCK component. Although having the same number of questions for each component does not guarantee to increase the reliability of the instrument, it could apparently provide security to researchers that use this questionnaire to evaluate PCK_{evo} of teachers. Another limitation is that the sample in pilot 2 is already small compared to other studies (e.g., Schmelzing et al., 2013; Jin et al., 2015; Kirschner et al., 2016; Park et al., 2018; Großschedl et al., 2019). This is explained by the difficulties getting teachers to answer the questionnaire. The reduced free time teachers must perform activities outside their teaching workload (Cofré et al. 2015) can explain this little availability of participation. This reduced teacher participation also prevented it from doing a factor analysis, which must be performed with more than a hundred people for an instrument of more than ten questions (Field, 2009).

This study presents a new instrument for measuring PCK which can be used along other methods such as a CoRe and PaP-eR (Loughran et al., 2004) or rubrics (Park & Oliver, 2008). For the study of PCK, different data sources are often triangulated. We are aware that an adequate capture of PCK in teachers is carried out "in action" and "on action"; however, it is important to note that declarative instruments have laid the foundations of what is known today about PCK in various topics. As researchers, we have advanced in complementary research of PCK "on action," based on the products of this pencil and paper instrument, finding consistency in the results (Bravo and Cofré 2016). This instrument contributes with one more source of reliable data (Chan & Hume, 2019). On the other hand, regarding the reliability of the results, the application of this instrument offers reliable data about PCK_{evo} in biology teachers. This can open a new line of investigation in the field of PCK_{evo}, where most studies have been performed with small samples (e.g., Bravo & Cofré 2016; Lucero et al., 2017). Lastly, this new instrument could facilitate the analysis of the relation between PCK_{evo} and other variables such as teacher subject matter knowledge or student performance (van Driel et al., 2014; Chan & Hume, 2019).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11165-022-10042-0.

Funding This study was supported by the International Cooperation program for attraction of advanced human capital, MEC 80180066 and by the Chilean National Fund for Scientific and Technologic Development (FONDECYT) grant number 1181801, both to HC.

References

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

AERA, Apa, & NCME [American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.] 2014 Standards for educational and psychological testing AERA Washington, DC

Alonzo, A. C., & Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching, 53*(8), 1259–1286.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561–573.

Aryadoust, S. V. (2009). Mapping Rasch-based measurement onto the argument-based validity framework. *Rasch Measurement Transactions, 23*(1), 1192–1193.

Baxter, J. A., & Lederman, N. G. (1999). Assessment and measurement of pedagogical content knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 147–161). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer.

Bond, T., & Fox, C. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Bravo, P., & Cofré, H. L. (2016). A new approach to capture and develop Biology Teachers' Pedagogical Content Knowledge through learning study: the case of human evolution. *International Journal of Science Education, 38*(16), 2500–2527.

Carlson, J., & Daehler, K. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–92). Sydney, Australia: Springer.

Chan, K., & Hume, A. (2019). Towards and consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In: Hume, A., R., Cooper y A., Borowski (Eds.). Repositioning pedagogical content knowledge in teachers' Knowledge for Teaching Science. Springer.

Coetzee, CM, Rollnick, E, & Gaigher (2020). Teaching electromagnetism for the first time: A case study of pre-service science teachers' enacted pedagogical content knowledge *Research in Science Education* https://doi.org/10.1007/s11165-020-09948-4

Cofré, H. L. C., Vergara, C., González-Weil, D., Santibáñez, G., Ahumada, M., Furman, M. E., et al. (2015). Science Teacher Education in South America: the case of Argentina, Chile, and Colombia. *Journal of Science Teacher Education, 26*(1), 45–63.

Cofré, H. L., Cuevas, E., & Becerra, B. (2017). The relationship between biology teachers' understanding of the nature of science and the understanding and acceptance of the theory of evolution. *International Journal of Science Education., 39*(16), 2243–2260.

Cofré, H. L., Jiménez, J., Santibáñez, D., & Vergara, C. (2016). Chilean pre-service and in-service teachers and undergraduate students' understandings of evolutionary theory. *Journal of Biological Education, 50*(1), 10–23.

Cofré, H. L., Santibáñez, D., Jiménez, J. P., Spotorno, A., Carmona, F., Navarrete, K., & C. A., Vergara,. (2018a). The effect of teaching the nature of science on students' acceptance and understanding of evolution: myth or reality? *Journal of Biological Education, 52*(3), 248–261.

Cofré, H.L., P. Núñez, D. Santibáñez, J. Pavez, C. A. Vergara (2018b). Theory, evidence, and examples about teaching nature of science and biology using history of science: A Chilean experience. In M. E. Brzezinski Prestes & C. C. Silva (Eds). T*eaching science with context: Historical, Philosophical, Sociological Approaches*. Springer.

Deniz, H., & Borgerding, L. (2018). *Evolution education around the globe*. Springer.

Ergönenç, J., Neuman, K., & Fischer, H. (2014). The impact of pedagogical content knowledge on cognitive activation and student learning. *Quality of Instruction in Physics*, 145–160.

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.

Fischer, H. E., Boone, W. J., & Neumann, K. (2014). Quantitative research designs and approaches. In N. Lederman & S. Abell (Eds.), *Handbook of research on science education* (Vol. 2, pp. 18–36). New York: Routledge.

Friedrichsen, P., Brown, L., & Schul, J. (2018). Project teach evolution: Preparing biology pre-service teachers to teach evolution in Missouri, USA. In: H. Deniz and L. Borgerding, ed., *Evolution Education Around the Globe*, 1st ed. 41–58.

Gess-Newsome, L. (2015). A model to teach professional knowledge and skill including PCK. In: A. Berry, P., Friedrichsen, & Loughran, J (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28–42). New York: Routledge.

Glaze, A., & Goldston, M (2015). US science teaching and learning of evolution: A critical review of literature 2000–2014. *Science Education* 99, 3, 500–518.

Großschedl, J., Welter, V., & Harms, U. (2019) A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching,* 56, 402–439.

Harms, U., & Reiss, M. J. (2019). *Evolution education re-considered*. Understanding what works: Springer.

Jin, H., Shin, H., Johnson, M., Kim, J., & Anderson, Ch. (2015). Developing learning progression-based teacher knowledge measures. *Journal of Research in Science Teaching, 52*(9), 1269–1295.

Jüttner, M., & Neuhaus, B. (2012). Development of items for a pedagogical content knowledge test based on empirical analysis of pupils' errors. *International Journal of Science Education, 34*(7), 1125–1143.

Kanter, D. E., & Konstantopoulos, S. (2010). The impact of a project-based science curriculum on minority student achievement, attitudes, and careers: The effects of teacher content and pedagogical content knowledge and inquiry-based practices. *Science Education, 94*(5), 855–887.

Kampourakis, K., & Zogza, V. (2007). Students' preconceptions about evolution: How accurate is the characterization as "lamarckian" when considering the history of evolutionary thought? *Science & Education, 16,* 393–422.

Kirschner, S., Borowski, A., Fischer, H., Gess-Newsome, J., & von Aufschaiter, C. (2016). Developing and evaluating a paper-and-pencil test to assess components of physics teachers' pedagogical content knowledge. *International Journal of Science Education, 38*(8), 1343–132.

Lederman, J. S., Lederman, N. G., Bartos, S. A., Bartels, S. L., Antink-Meyer, A., & Schwartz, R. S. (2014). Meaningful assessment of learners' understandings about scientific inquiry—The views about scientific inquiry (VASI) questionnaire. *Journal of Research in Science Education, 51*(1), 65–83.

Loughran, J. J., Berry, A. K., & Mulhall, P. (2012). *Understanding and Developing Science Teachers' Pedagogical Content Knowledge*. The Netherlands: Sense Publishers.

Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching, 41*(4), 370–391.

Loughran, J., Mulhall, P., & Berry, A. (2008) Exploring pedagogical content knowledge in science teacher education. *International Journal of Science Education,* 30(3), 1301–1320

Lucero, M., Petrosino, A., & Delgado, C. (2017). Exploring the relationship between secondary science teachers' subject matter knowledge and knowledge of student conceptions while teaching evolution by natural selection. *Journal of Research in Science Teaching, 54*(2), 219–246.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature sources and development and pedagogical content knowledge for science teaching. In: Gess- Newome & N.G.Lederman (Eds.). *Examining* *pedagogical content knowledge: The construct and its implications for science education* (pp. 95–132). Dordrecht: Kluwer Academic.

Mineduc (2015). Bases Curriculares 7° básico a 2° medio. Santiago: Ministerio de Educación [National Curriculum, grade seventh to tenth].

Mineduc. (2019). Bases Curriculares 3° y 4° medio. Santiago: Ministerio de Educación. [National Curriculum, grade eleventh to twelfth].

Nehm, R. H. (2018). Evolution. In: Teaching biology in Schools. *Global research, issues, and trens*. In: K. Kampourakis y M. Reiss (Eds.).164 – 177. Routledge.

Nehm, R. H., & Schonfeld, I. S. (2007). Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching ofevolution in schools? *Journal of Science Teacher Education, 18*(5), 699–723.

Padilla, K., Ponce de León, A., Rembado, F. M., & Garritz, A. (2008). Undergraduate professors' pedagogical content knowledge: the case of 'amount of substance.' *International Journal of Science Education, 30*(10), 1389–1404.

Park, S., & Oliver, J. (2008). Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38*(3), 261–284.

Park, S., Jang, J.-Y., Chen, Y.-C., & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching?: Evidence from an empirical study. *Research in Science Education, 41*(2), 245–260.

Park, S., & Suh, JK. (2015). From protraying toward assessing PCK: Drivers, dilemas, and directions for future research. In A. Berry, P. Friedrichsen, y J. Loughran (Eds.), *Re-examining Pedagogical Content Knowledge in Science Education* (pp. 104–119). Routledge.

Park, S., Suh, J., & Seo, K. (2018). Development and validation of measures of secondary science teachers' PCK for teaching photosynthesis. *Research in Science Education, 48,* 549–573.

Parraguez, C., P., Núñez, D. Krüger & Cofré, H. L. (2022). Describing changes in student thinking about evolution in response to instruction: the case of a group of Chilean ninth-grade students. *Journal of Biological Education*.

Reynolds, M., & Park, S. (2021). Examining the relationship between the educative teacher performance assessment and preservice teachers' pedagogical content knowledge. *Journal of Research in Science Teaching, 58*(5), 721–748.

Robbins, J., & Roy, P. (2007). The natural selection: Identifying & correcting non-science student preconceptions through an inquiry-based, critical approach to evolution. *The American Biology Teacher, 69*(8), 460–466.

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*, 1–36.

Romine, W., Barnett, E., Friedrichsen, P., & Sickel, A. (2014). Development andevaluation of a model for secondary evolution educators' professional development needs. *Evolution: Education and Outreach, 7*(1), 27. doi: https://doi.org/10.1186/s12052-014-0027-y

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. Z. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching, 48*(2), 117–148.

Schmelzing, S., van Driel, J., Jüttner, M., Brandenbusch, S., Sandmann, A., & Neuhaus, B. (2013). Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the "cardiovascular system" *International Journal of Science and Mathematics Education, 11*(6), 1369–1390.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15,* 4–14.

Sickel, A., & Friedrichsen, P. (2013). Examining the evolution education literature with a focus on teacher: major findings, goals for teacher preparation, and directionsfor future research. *Education and Outreach, 6*(23), 2–15.

Sickel, A., & Friedrichsen, P. (2018). Using multiple lenses to examine the development of beginning biology teachers' pedagogical content knowledge for teaching natural selection simulations. *Research in Science Education, 48,* 29–50.

Taber, K. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48,* 1273–1296.

van Dijk, E. M. (2009). Teachers' views on understanding evolutionary theory: A PCK study in the framework of the ERTE-model. *Teaching and Teacher Education, 25,* 259–267.

van Driel, J., Berry, A., & Meirink, J. (2014). Research on science teacher knowledge. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (Vol. 2, pp. 848–870). New York, NY: Routledge.
Williams, M., Sadler, T., & Friedrichsen, P. (2018). Modeling the emergence of antibiotic resistance in bacterial populations. *The American Biology Teacher, 80*(3), 210–216.
Zhou, S., Wang, Y., & Zhang, C. (2016). Pre-service science teachers' PCK: Inconsistency of preservice teachers' predictions and student learning difficulties in Newton's third law. *EURASIA Journal of Mathematics, Science and Technology Education, 12*(3), 373–385.

Authors and Affiliations

Beatriz Becerra[1] · Paola Núñez[1] · Claudia Vergara[2] · David Santibáñez[3] · Dirk Krüger[4] · Hernán Cofré[1,5]

Beatriz Becerra
bbecerraolguin@gmail.com

Paola Núñez
paola.nunez@pucv.cl

Claudia Vergara
clvergara@uahurtado.cl

David Santibáñez
dsantibanez@uft.cl

Dirk Krüger
dirk.krueger@fu-berlin.de

1 Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

2 Universidad Alberto Hurtado, Santiago, Chile

3 Universidad Finis Terrae, Santiago, Chile

4 Freie Universität Berlin, Berlin, Germany

5 Instituto de Biología, P, Universidad Católica de Valparaíso, Av. Universidad 330, Curauma, Valparaíso, Chile