Review Article

Received: July, 20.2020.
Revised: August, 10.2020.
Accepted: August, 19.2020.

UDK:
37.091.12
159.923
doi: 10.5937/IJCRSEE2002121C

Peer Assessment of Teacher Performance. What Works in Teacher Education?

Valeria M. Cabello[1], Keith J. Topping[2]*

[1]Pontificia Universidad Católica de Chile, Faculty of Education and Research Center for Integrated Disaster Risk Management (CIGIDEN), Santiago, Chile, e-mail: vmcabello@uc.cl
[2]University of Dundee, School of Education and Social Work, Dundee, Scotland, e-mail: k.j.topping@dundee.ac.uk

Abstract: Peer assessment is increasingly used in schools and higher education, especially in health education. However, there remains insufficient evidence that peer assessment conditions are beneficial for teacher education. In this article, empirical research literature on peer assessment of pre-service teaching performance are reviewed. The articles were from the ERIC and Scopus databases, from 2002 to 2020. Only fifteen studies met the selection criteria described herein. The studies differed in the type of assessment used but converged toward the conclusion that incorporating peer assessment into different stages of teacher education was appropriate and worthwhile. We discuss the theoretical perspectives on why peer assessment might work in teacher education, pointing out practical implications for decision-makers in this field. Finally, recommendations and constraints for researching and implementing peer assessment are discussed from the perspective of innovation within pre-service teacher education.

Keywords: deviant online behavior, minors, young adults, social networks, risk factors, vulnerability, resources, risk assessment.

Introduction

In the last decade, an increasing interest in peer-learning and peer-assessment (PA) in higher education has occurred, especially in health profession areas (Arnold et al., 2007). However, current evidence does not necessarily support its worthiness (McNulty, 2019). In the field of teacher education, teaching in front of peers, is a quite common practice for diverse purposes (Abdulwahed, 2011; Amobi, and Irwin, 2009), with positive effects in self-efficacy beliefs (d'Alessio, 2018). Despite the relevance of an authentic assessment process to monitor the development of pre-service teacher competences, those which rely on peers as a source of information and collaborative learning are still scarce (Charalambous, Hill and Ball, 2011, Ratminingsih, Artini and Padmadewi, 2017). Indeed, the evidence supporting PA is less robust than other types of assessment in the educational field (Li et al. 2020), which constitutes a research problem, not only for teacher educators bur for decision-makers in this field.

The studies which form the subject of this review, focus on pre-service teachers (PST) teaching performance – rather than on more typical peer-assessed tasks such as written assignments (e.g. Tsai, Lin and Yuan, 2002). Although other prior reviews of PA exist (e.g., Gielen, Dochy, and Onghena, 2010; Li et al. 2020; van Zundert et al. (2010)), these are not focused on teaching performance nor were developed for teacher education. Hence, the present review analyzes a more specific area. Its purpose is to orient stakeholders to critical aspects of the design of PA in teacher education which are empirically based and focused on issues that require further research in order to gain sufficient understanding.

We start by defining PA, its benefits and disadvantages, followed by the essential organizational features and its implementation. After this, the reviewed studies are discussed in terms of outline trends and connection of principles for PA in pre-service teachers and practical implications for working PA in teacher education.

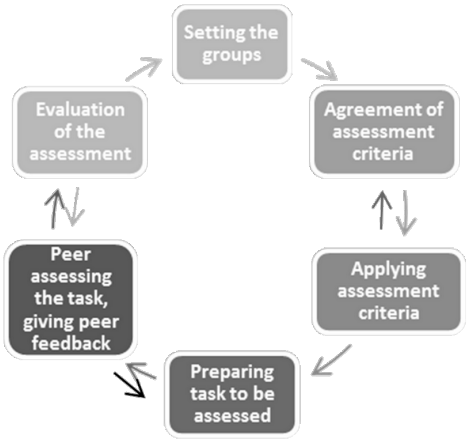
*Corresponding author: vmcabello@uc.cl

Definition and Types of Peer Assessment

Peer assessment (PA) is a form of evaluation that is designed for enhancing learning (van Gennip, Segers, and Tillema, 2009). Thus, apart from serving an evaluative function, it offers a learning opportunity (Bunch, Aguirre, and Tellez, 2009). PA is understood as "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equalstatus learners" (Topping, 2009, p.20). PA by itself or as a complement to other types of assessments has broad areas of application (Gielen, Dochy and Onghena, 2010; Li et al., 2020; van Zundert et al., 2010). The purpose of PA on performance is to help learners make judgments about structured tasks and provide their impressions to peers. PA processes include judging whether specific actions are performed, their quality and suitability for a purpose (Norcini, 2003).

In its nature, PA is a social process, in which one of the essential components is the feedback given to and received from others (Sluijsmans and Prins, 2006; van Genip et al., 2009). Peer feedback is usually reciprocal between assessor and the individual assessed. It can be delivered face-to-face or remotely, verbally or in a written form, immediately or delayed. It can have an affirmative, corrective or suggestive orientation, and reduce errors if received thoughtfully. Useful feedback requires understanding the assessment goals and criteria, and the ability to judge the relationship of the specific performance to these goals (Topping, 2010). PA as an assessment method can be summative, formative or both (Gielen et al., 2010). Formative assessment involves participants helping each other to identify their strengths, weaknesses, and target areas for remedial action, aiming to develop metacognitive skills for future performance (van Gennip et al., 2009). Otherwise, a summative assessment gives feedback often when it is too late to affect the production of the present task, although it may affect the production of future tasks (Topping, 2010). Figure 1 shows the essential organizational features of PA in education.

Figure 1.
Typical PA implementation in education



In this process, working in small groups and trying to avoid close friends or enemies/adversaries has been suggested to facilitate group involvement in the assessment (Topping, 2010). The person in charge helps the group to agree on the assessment criteria that will be used in the PA. This is followed by the exemplary application of the criteria to past cases/evidences or representative examples of the task. Using anonymous examples is recommended to avoid anxiety (Sluijsmanset all., 2003). The application of the assessment criteria is again discussed and clarified. The participants are then encouraged to prepare the performance of a task knowing that it will be peer-assessed. Furthermore, the PA is conducted using the agreed assessment criteria. Kilic and Cakan (2007) recommended between three and five assessors in a set, so that participants (assessees - those being assessed) can balance feedback from different peers, which helps enhance PA reliability. Peer feedback is given to offer the assessee the possibility of improving the performance. The quality of subsequent performances is therefore relevant. The person in charge should evaluate the process to encourage accuracy in applying the criteria and giving peer feedback. Finally, reworking the task in the light of peer feedback is essential, to promote a sense of agency in the assessee.

The cycle may be repeated with the same or different groups to complement the feedback. Wen and Tsai (2008) suggested three rounds of PA. Nevertheless, it is essential not to overload the participants with too many loops, because the benefit of gaining more feedback has a cost in terms of time, which might become a disadvantage.

Benefits and Disadvantages

van Zundert et al. (2010) indicated that most studies on Peer Assessment (PA) had shown benefits, but disadvantages can also be identified. Nevertheless, most of these were identified on researchers' opinion more than on evidence. This poses the need for reviewing empirical studies on the topic.

Benefits of PA in education:

- Increasing reflection upon and generaliation of learning to new situations (Kim, 2009; Ratminingsih, Artini and Padmadewi, 2017)
- Encouraging students' self-regulation and self-awareness (Asghar, 2009)
- Improving the students' dispositions related to being assessed (Ratminingsih, Artini and Padmadewi, 2017)
- Developing students' self-concept (Duran and Monereo, 2008).

Furthermore, PA can save teaching time because of the more immediate and individualized feedback from peers (Sun et al., 2019). Nonetheless, this saving is not often achieved in the short run because the implementation of good quality PA requires a period for organization, training, and monitoring (Falchikov, 2001).

Disadvantages of PA in education:

One of the reported difficulties of PA is the amount of time required for the organizers and the participants (Okhremtchouk et al., 2009). To help with this problem, PA should be integrated into the curriculum (Kilic and Cakan, 2007; Strijbos and Sluijsmans, 2010). Likewise, initial reluctance and anxiety to participate is quite frequent (Arnold et al., 2005). Assessors beginning with positive feedback to the assessee could reduce this and improve subsequent acceptance of more critical feedback (Topping, 2010). Also, discussion, negotiation, and joint construction of assessment criteria with concrete exemplary material before PA might be worthy (MacArthur, Schwartz, and Graham, 1991). Indeed, performing in front of peers might be less stressful than doing it for the first time in front of teachers (Britton and Anderson, 2010).

Another issue is the reliability of PA, because friendships, popularity, enmity, perception of criticism as socially uncomfortable, or the trend to assign average scores can all be affected respectively. Nonetheless, using performance checklists or rubrics, extensive exemplification and careful monitoring of the PA process can increase reliability (Topping, 2009). Ensuring validity, reliability, and fairness of the measures is a more critical issue when assessments are used to make high-stakes decisions (Sandholtz and Shea, 2012). Hence, having clear assessment criteria (Sluijsmans and Prins, 2006), more than one assessor and anonymity between assessors and assessee might help in both summative and formative assessment (Kilic and Cakan, 2007; Vickerman, 2009). This is especially relevant in professional learning contexts. According to Gielen et al. (2011), PA between teachers can be an excellent way to assess teaching skills and also to improve them, but only if all different peer opinions enrich the assessment and if the assessment criteria are clear for all teachers. Although the benefits and disadvantages of PA are known, there still exists a gap in knowledge with regards to which characteristics of PA are supported by evidence within the context of teacher performance. Thus, this review is required.

Peer Assessment of Performance in Teacher Education

One of the goals for pre-service teacher education is to prepare student teachers to critically reflect on their conceptions about teaching, their practice and their peers' practice (Amobi and Irwin, 2009; Feiman-Nemser, 2008). Peer Assessment (PA) in teacher education is a strategy for helping learners to examine their progress in teaching (Sluijsmans and Prins, 2006) and be familiar with it before teaching in classrooms (Wen and Tsai, 2008). Teaching practice is an exercise used to expose student teachers to the practical aspects of teaching (Oluwatayo and Adebule, 2012). Although teaching practice is essential in teacher education (Jian, Odell, and Schwille, 2008; Oluwatayo and Adebule, 2012), attempts to improve it through early assessment are not widely reported (Charalambous, Hill and Ball, 2011).

Moreover, if there is a lack of performance assessment during pre-service education, student teachers do not know if they possess the required classroom teaching skills, nor the quality criteria used to measure their performance or how to improve it (Oluwatayo and Adebule, 2012). Likewise, considering

the difficult shift from the student role to the teacher role without adequate support in the transition process (Jian, Odell, and Schwille, 2008), the lack of clear and agreed recognition of teaching competences is a problem. However, early experiences in the classroom – for example, by student teachers observing an expert teacher and slowly taking more teaching actions - might help in this (Hume, 2012), also developing skills to critically observe teaching practice (Sonmez and Can, 2010). Furthermore, PA in early teaching stages enhances student teachers' ability to analyze and reflect on their practice (Harford and MacRuairc, 2008), i.e., it makes them more analytical when appraising the teaching performance (Sluijsmans et al., 2004), and more able to bridge the gap between their conceptions and practice (Ostroskyet al., 2013). Despite these potentialities of PA, there is still uncertainty on which specifically are supported by recent evidence and what kind of interventions work for teacher education. These particular questions constitute the problem the present review will help to solve. PA studies in other areas of higher education have contradictory conclusions, thus, there is a gap in empirical research oriented to aspects of PA that might be more efficient in teacher education and assure its benefits can overcome the respective difficulties (Li et al. 2020).

Materials and methods

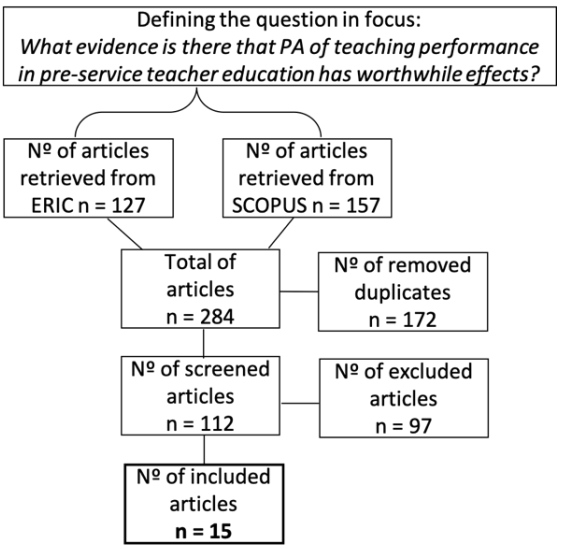
This article reviews empirical studies from 2002 to 2020 in PA of teaching performance in pre-service teacher education to inform researchers and decision-makers, answering the research question 1: (RQ1) What evidence is there that PA of teaching performance in pre-service teacher education has worthwhile effects?

Moreover, in a broader scope, this article seeks to answer the stated questions using the antecedents and the systematic review, the research question 2: (RQ2) What are the theoretical underpinnings of PA of teaching performance in pre-service teacher education?

Review of Studies

A systematic review was used to approach this work following the steps recommended by Cook and West (2012); 1. Defining the question in focus, 2. Identifying information sources – we decided to use two main educational databases; ERIC and Scopus, 3. Searching for eligible studies with defined search terms -we used 'peer assessment' + teach* + performance*. 4. Defining inclusion criteria –English language, articles only with open access. 5. Defining exclusion criteria – articles not based on empirical research. 6. Defining data abstraction elements – we removed duplicates based on the title. 7. Analyze and synthesize – we screened the articles and excluded those not centered on teaching performance in pre-service teachers. Following these steps, we arrived to include fifteen articles (Figure 2).

Figure 2.
Schema of steps in the review process



The fifteen studies included in the research review were conducted in different places, with diverse conditions and varied aims, as summarized in Table 1. The nature of the studies allowed grouping of

them into categories which are described in the following section. Studies focused on PA of teaching performance for its improvement are reviewed first (A); studies centered on pre-service teachers', specifically their development of assessment skills (B), later.

Table 1.
Overview of studies

Study	Category	Declared aim	Author(s) and year	Context/place	Participants
1	A	To foster a community of practice to address the challenges of teaching	Harford & MacRuairc (2008)	Ireland. PA of 10-minute clip of a real lesson during the first teaching practice	20 final-year PST of varied subjects
2	A	To assess teaching skills by peers and instructor	Kilic & Cakan (2007)	Turkey. Microteaching in small groups	122 third-year science PST
3	A	To examine the relationship of students' gender with PA	Oren (2012)	Turkey. Presentation skills for science teaching	203 science PST
4	A	To examine PA, self-assessment and teacher assessment	Kiliç (2016)	Turkey. Presentation performances muti-assessed	15 second-year PST
5	A	To investigate student teachers' thinking and performance to deliver instructional explanations	Charalambous, Hill & Ball (2011)	The United States of America. Microteaching in a simulated classroom	16 final-year primary PST
6	A	To explore the development of explaining skills fostered by peers	Authors (2018)	Chile. Microteaching in small groups	20 fourth-year science PST
7	A	To explore the effects of anonymity in PA within a Facebook-based app.	Lin (2018)	Taiwan. PA of two groups based on microteaching performance	32 PST of adult learners' education
8	A	To help PST to enhance their self-efficacy beliefs.	d'Alessio (2018)	The United States of America. PA of microteaching	433 primary PST
9	A	To investigate the contribution of PA in fieldwork preparation	Al-Barakat & Al-Hassan (2009)	Jordan. PA of practicum experience (45-minute lessons)	72 early childhood PST in the last year
10	A	To investigate the use of peer assessment dialogue as an assessment for learning tool	Eather, Riley, Miller & Jones (2017)	Australia. PA of practical tutorial activities oriented to build skills in peer assessment and feedback	36 physical education PST
11	B	To develop the ability to define performance criteria	Sluijsmans et al. (2004)	The Netherlands. PST enrolled in a teaching design course	93 second-year primary PST
12	B	To investigate assessee's role on metacognitive awareness, performance and motivation towards PA	Kim (2009)	Korea. PST enrolled in an Educational Technology course	82 secondary mathematics PST
13	B	To investigate the impact of PA in assessment skills: how to give feedback and write an assessment report	Sluijsmans et al. (2003)	The Netherlands. PA of writing reflection papers	110 first-year mathematics PST
14	B	To identify PST attitudes towards PA and assessment skills	Seifert & Feliks (2019)	Israel. PA of peers' assignments and performance	300 bachelor's and master's degree PST
15	B	To compare the students' perceptions of two PA instructional designs	Mercader, Ion & Díaz Vicario (2020)	Spain. PA of enacted case studies in two instructional designs	556 second-year PST

(A) Studies based on peer assessment of teaching performance

Harford and MacRuairc (2008) underlined the relevance of formative PA and feedback. In their study, the student teachers gradually moved their focus of analysis towards more meaningful reflection. They deconstructed the practice of their peers more critically and analytically. The researchers used focus groups, the results of which suggested that student teachers had developed their reflective skills. Moreover, pre-service teachers felt able to transfer the good practice observed in their work, evaluating the project as a powerful mechanism for conducting self-review. They valued informal and formative feedback, remarking that formal assessment would have reduced their engagement with the process and

the quality of the reflective dialogue. In this study, intervention and research were conducted by the same researchers. The results were based on a qualitative analysis of self-reports, so its reliability is open to challenge.

In the study of Kilic and Cakan (2007), the peer assessors and instructor evaluated pre-service science teachers' content and teaching knowledge, the teaching and learning process, class management and communication, as presented in a microteaching episode. They used a form with a 5-point Likert-type scale between very good and very poor. Peer scores were considerably higher than the instructor's, but the two scores significantly correlated. Improved correlations were then found in the second attempt at PA. They showed the number of peer assessors needed to be around five to sustain acceptable reliability. This study did not determine if the assessed performance improved.

Oren (2012) studied the influences of participant gender in scores from peers, self and instructor on communication skills for teaching science during a 10-week semester. In this study, female participants obtained significantly higher mean scores than males in all score types. However, as this study did not count with a baseline measurement, we do not know if female pre-service teachers had better skills or they benefited more from the course. Similarly, Kiliç (2016) conducted multiple assessments by peers, self and instructor. The results showed PA was significantly higher than the other types of evaluations based on their scores. PA was perceived as an enhancer of successful performance and higher confidence for presenting lessons between student teachers. Although this assertion is based on participants perceptions and not supported by a more objective indicator, it is a positive signal for strengthening PST's self-confidence.

Charalambous, Hill and Ball (2011) investigated mathematics pre-service teachers' thinking and performance when delivering instructional explanations. The 20 participants co-constructed a list of criteria for determining the quality of instructional explanations, giving a performed example in a simulation of a lesson. The peers completed reflection cards and shared their comments with the performers. Through a case analysis of four participants, the researchers concluded that their performance could grow to vary degrees after PA and reflection.

The study of Cabello and Topping (2018) had a within-subject repeated-measures design. The 20 student teachers received training in PA and constructed assessment criteria. They peer-assessed microteaching in two rounds through a rubric. The PST significantly increased their microteaching performance after PA, with a reasonable effect size (d=1.4). This improvement was maintained and transferred into real-life teaching when followed up. Nonetheless, the participants' self-selection and the limited sample size affected the generalizability of the results.

In the experimental study of Lin (2018), online PA was carried out in a Facebook-based learning application, looking for effects of anonymity on affective, cognitive and metacognitive peer feedback. The student teachers were randomly assigned to write feedback to five peers' microteaching videoed performance in an anonymous or identifiable condition. The author applied a 6-point Likert scale for the perceived learning, fairness of peer feedback and attitude toward the online PA system. The anonymous group gave more cognitive feedback and the identifiable group more affective and metacognitive. In the role of assessees, the group, when in an unidentified condition had a better attitude towards the system. However, they perceived peer comments as less fair than those of the participants in the identifiable group. The study argues that the cognitive and pedagogical benefits of anonymity in online PA are well demonstrated, although, the data analysis only allows inferences on their perceptions.

d'Alessio (2018) conducted a study to help student teachers to build self-efficacy beliefs. The student-teachers performed microteaching, self, and peer assessment of the events, which was analyzed using a rubric. The quality of microteaching based on PA and mastery of the content had the most significant influence on participants' self-efficacy beliefs in this study.

Al-Barakat and Al-Hassan (2009) explored early childhood student teachers' preparation in various applied contexts. They received extensive training on PA reviewing videos. Then each PST was observed once a week by 4-5 peers over ten weeks in a 45-minute lesson. The classroom observations were discussed in a group, and feedback was given. Using interviews, they found PA developed participants' classroom performance, especially on competencies such as designing objectives, activities, teaching strategies, interaction, students' assessment and classroom management. Moreover, they found PA helped student teachers to form a set of criteria for judgement on classroom performance and improvement of self-confidence within the assessment.

Similarly, the use of peer dialogue assessment was researched by Eather et al., (2017) in physical education PST. They found significant improvements in perceived teaching confidence and competence, and teaching self-efficacy based on self-reports. In both studies, the same researchers conducted the interviews and the course, making the validity more open to challenge.

(B) Studies based on peer assessment of performance as a skill

In the study of Sluijsmans et al., (2004), the participants were randomly assigned to similar-sized control (n=47) and experimental (n=46) groups. The experimental group was trained in PA while the control group was not. The student-teachers defined a set of assessment criteria for designing a creative lesson plan. The study used PA forms, questionnaires and interviews. The researchers found the experimental group was more capable in applying the criteria than the control group, the experimental group also used the criteria more often and felt more able to assess after PA than before PA. Still, there was no significant effect of the training in PA skills on performance. Even so, the researchers concluded that PA skills could be successfully trained, but the value of the training is uncertain if no impact was found on their resultant performance.

Kim (2009) used a metacognitive awareness questionnaire and a motivation survey. PST's performance was measured in an assignment to create a concept map on instructional design. All the participants submitted their tasks for peer feedback and tutor marks. After receiving feedback, the participants were randomly assigned to an experimental condition that received a back-feedback opportunity (n=40) and a control condition without back-feedback (n=42). Back-feedback consisted of giving their opinion on the feedback, enabling reflection on peer feedback. After revision, PST resubmitted their concept maps and completed the metacognitive awareness questionnaire and survey. The experimental group subsequently showed higher metacognitive awareness, better performance and better attitudes towards PA than the control group. The researcher did not report the size of effect of the improvement, or the correlations between peer and tutor marks. Thus, the study seems unspecific in the manner which is published.

The study of Sluijsmans et al., (2003) had a within-subject repeated-measures design. Questionnaires, PA forms and the student teachers' reflections were used to assess written reflection papers. PST received several training sessions on assessment skills, how to give feedback and write assessment reports. They agreed nineteen assessment criteria (i.e. self-criticism, work field experiences, personal expectations, etc.) and marked the peers' task and wrote their written reflections in a virtual learning environment. The instructor also marked the reflection papers. The researchers concluded there was significant progress for most variables studied: the participants used the assessment criteria better; their feedback was better, and their assessment reports were more structured. Likewise, they wrote better reflection papers, based on the instructor's marks; however, the effect size of the advance was not reported. Moreover, the design of the study does not allow PA to be related causally to better performance because a comparison group was not incorporated.

Seifert and Feliks (2019) studied attitudes concerning self-assessment and anonymous PA to improve PST assessment skills. The participants assessed several products and noted they benefitted from the process and developed good attitudes to PA. Mercader, Ion and Díaz-Vicario (2020) used different instructional designs to guide PA. PST perceived that long-term mediations, two rounds of PA and giving feedback were the most useful. Both studies had quantitative analysis and linked them with PST's perceptions.

Results and Discussions

Here the research findings of this review are summarized, then interpreted. After this, they are discussed in terms of practical implications for teaching and future research.

The analysis of the studies showed some similarities and several differences. Firstly, they were conducted from the early to the final years of teacher education, but the trend in similarity seemed to be stronger towards the latter years. The participants were from a wide range of subjects, with a slight tendency towards science and mathematics. The sample size varied from 16 to 556. The performances assessed were based on teaching skills (Al-Barakat and Al-Hassan, 2009; Charalambous, Hill and Ball, 2011; Charalambous, Hill and Ball, 2011; Kilic and Cakan, 2007; Lin, 2018), assessment skills (Kim, 2009; Mercader et al., 2020; Seifert and Feliks, 2019; Sluijsmans et al., 2004) and a combination of teaching practice with the development of peer assessment as a skill (Cabello and Topping, 2018; Sluijsmans et al., 2004). The purposes of PA were summative (Kilic and Cakan, 2007; Sluijsmans et al., 2004), formative (Al-Barakat and Al-Hassan, 2009; Cabello and Topping, 2018; Charalambous, Hill and Ball, 2011; Harford and MacRuairc, 2008; Kim, 2009; Lin, 2018) or both (Sluijsmans et al., 2003). Feedback was face-to-face in most of the studies, but d'Alessio (2018), Lin (2018), Mercader et al. (2020) and Sluijsmans et al. (2003) who used an online platform.

The studies also differed depending on the assessment criteria used for PA. Kilic and Cakan (2007),

Lin (2018) and Al-Barakat and Al-Hassan (2009) used criteria defined by the staff member, while Cabello and Topping (2018), Charalambous, Hill and Ball (2011), Sluijsmans et al. (2004) and Sluijsmans et al. (2003) agreed on the criteria between the participants. Harford and MacRuairc (2008) and Kim (2009) did not use structured criteria to guide the PA processes, although they found an improvement in awareness about teaching practice. Only some studies reported a guided training received by the participants (Al-Barakat and Al-Hassan, 2009; Cabello and Topping, 2018; Kiliç, 2016; Mercader et al., 2020; Sluijsmans et al., 2003).

Most of the studies reported benefits of PA on the performance assessed. However, the study of Sluijsmans et al. (2004) did not show an effect, Kilic and Cakan (2007), Lin (2018) and d'Alessio (2018) did not mention it. Nonetheless, two studies used a repeated measures design and reported a measurable improvement in the performance (Cabello and Topping, 2018; Sluijsmans et al., 2003). They found a significant and robust advance in participants' performance after PA - despite the vast difference in their sample sizes (20 vs 110), the type of performance assessed and the length of PA training. Even so, without a comparison group, establishing the cause of progress is open to challenge.

Lastly, only three studies were found with an experimental design, which allows linking the results obtained through PA to outcomes in a causal relation. Sluijsmans et al. (2004) did not find statistical differences that supported an effect on the assessed performance. Kim (2009) reported significant differences, but not the effect size of the improvement. Lin (2018) related one condition of PA -anonymity with the peer feedback and perceived learning, fairness and PA attitudes: not with effects on the teaching performance.

Most of the studies had quantitative measurements, but some of these were questionable in their reliability or accuracy, perhaps due to the limiting conditions of researching in higher education contexts, such as lack of possibilities for randomizing groups or having external evaluations on student teachers. This finding supports the view that research in PA needs more systematic work (Li et al. 2020).

Moreover, the short-term parameters of the studies reviewed here must be considered. Similarly, only two of them dealt with the transference of teaching practice to real contexts (Cabello and Topping, 2018, Al-Barakat and Al-Hassan, 2009). Thus, generalization and maintenance of the effects of PA in teacher education into work contexts are not robustly supported by evidence so far.

Theoretical Basis of Peer Assessment in Teacher Education

Some authors have presented theories or ideas to help understand the role of PA with respect to performance in teacher education. For instance, negotiation of meaning is a construct that might explain the possible success of formative PA in teacher education, primarily when the student teachers design the assessment criteria (Al-Barakat and Al-Hassan, 2009; Stiggins, 1991), which has been empirically tested by Sluijsmans et al. (2004). We strongly believe the crucial benefit of this is the construction of a third space in between common knowledge and teaching knowledge, where student teachers can jointly redefine what elements constitute good teaching. This construction is triggered with the interaction with peers, through assessment reflection and discussion, as discussed by Ratminingsi, Artini and Padmadewi (2017).

Additionally, PA provides students with skills to form judgments about what constitutes high-quality work (Cabello and Topping, 2018), and this challenges their conceptions about good teaching (d'Alessio, 2018). Moreover, PA within teacher education could enhance self-regulated learning, by giving student teachers the opportunity to talk about their decisions, beliefs and practices (Vermunt and Endedijk, 2011). This might lead to them gradually becoming the owners of their learning processes when they actively construct new ideas in interaction with peers (Ratminingsih, Artini and Padmadewi, 2017). However, self-regulation of learning is a necessary but not entirely sufficient condition (in of itself) to develop pre-service teachers' conceptions and skills (Vermunt and Endedijk, 2011). Cabello (2017) argued that changes in student teachers' conceptions and practice during PA might occur on the basis of two cognitive mechanisms: projection and reflection. The assessee's performance reflects what student teachers in the assessor role would do in a similar situation. The assessors identify themselves with this practice because a peer, who shares experiences performs it directly to them. Thus, the assessors project their own possible decisions and practice on the assessee's performance. Furthermore, discussing the assessment criteria and using them to analyze their practice gives a shared space for critical reflection on typical teaching performance, possibilities and understanding which might enhance changes in their conceptions with consequences towards their practice.

Critical reflection itself is required for making reliable judgments about peers' work because a comparison of peers' performance against teaching performance criteria is required (Ratminingsih, Artini

and Padmadewi, 2017; Sluijsmans and Prins, 2006). Likewise, critical reflection can also develop self-assessment skills and help student teachers improve their practice. As Stiggins (1991, p. 38) stated, "once students internalize performance criteria and see how those criteria come into play in their own and each other's performance, students often become better performers". This might be one of the reasons for the positive results obtained by Kiliç (2016). Thus, PA is understood as a cognitive and social activity to enhance student teacher professionalism. In relation to this point, Lin (2018) states that PA in teacher education is an integral part of the learning process. We extend the argument for PST, based on the idea that discussing assessment forms or rubrics with peer assessors is crucial (Kilic and Cakan, 2007).

Considering that PA is an activity that all professionals may expect to experience at different times of their professional life, implementing PA at university seems to reflect demands for transferable skills. PA can help student teachers face changing teaching environments by giving them not only an active role in the detection and remediation of their weaknesses (Inoue, 2009) but also in the development of their communication skills and collaboration (Sluijsmans and Prins, 2006). Even so, collaboration is not the only explanation of why PA might be efficient in teacher education. The internalization of assessment criteria for enacting good quality performance (Stiggins, 1991) is an underlying principle, which might function as an enhancer of self-regulation in teaching practices (Vermunt and Endedijk, 2011), from a cognitive perspective (Kollar and Fischer, 2010).

Practical implications

Although most of the studies reviewed supported the feasibility of PA in teacher education, some studies were questionable in their methods. Disadvantages of PA can appear during its implementation. Most of these can be avoided - i.e. anxiety - but some are unavoidable, such as the time required for conducting good quality PA. It is vital that stakeholders who decide to embed PA in teacher education take careful actions to minimize the disadvantages. For instance, arranging the PA groups to avoid adversarial instances, assuring anonymity and applying the assessment criteria with others' performance first, have all been suggested (Lin, 2018).

The incorporation of PA in pre-service teacher education helps the diagnosis of competences (Sluijsmans and Prins, 2006) and understanding of how effective the teacher education program is being or has been (Pecheone and Chung, 2006). For instance, PA can reveal certain shortcomings of the students within the program, such as students lacking skills to analyze teaching practice or to give feedback (Sonmez and Can, 2010), or informing about areas of strength and weakness (Darling-Hammond, 2006). This is a form of accountability in teacher education programs. Thus, PA could bridge the gap between instruction and assessment, monitoring student teachers' progression and helping them to improve. This point might be interpreted as the potential of PA in initial teacher education, as a learning, teaching, diagnosis and intervention tool at the same time, and consequently, a cost-effective innovation in teacher education. However, to state this argument, more evidence is needed to support the ideas.

Furthermore, it would be of interest and benefit to investigate whether establishing a continuum of PA can have an impact on professional teachers' skills. For instance, comparing programs that systematically use PA from the early years of teacher education with others that only use them in the final years or even in only teaching placements. This is under the assumption that it might be advantageous to include PA when student teachers already have some practical experience, so they can use it to further strengthen their teaching competences. Preparing student teachers from the early stages of teacher education in PA skills could be an option for creating a culture of formative assessment in teacher education. Nonetheless, more robust evidence is needed to support advances in their teaching due to PA than is currently available.

The possible effect of PA on the professional identity of PST is also an exciting field to explore. It is known that the first years of teaching are essential to model teaching practice, and in this period, new teachers receive several influences from colleagues and mentors (Day, 2008). Thus, exploring the movements of identity when future teachers take on the judgement of the practice of others -and themselves- could adequately a complementary systemic for preparing teachers for self-regulated learning and lowering their dependence on others' judgements of the quality of their practices. Further research may also explore the influence of PA between peers on the self-concept of early student teachers.

van Zundert et al. (2010) and Ratminingsi, Artini and Padmadewi (2017) indicated that there is a lack of research into PA as an integral part of pre-service teacher education. Thus, the extent to which widespread PA would carry different benefits in embedded versus focal interventions is a question emerging from this review.

Of course, this review has some limitations, regarding for instance, only articles written in English were analyzed. This might unintentionally have restricted the broad view of PA. However, the reports

mentioned covered several countries and not exclusively in English speaking contexts. Thus, this review is still broad in its scope in terms of its research locations. Nevertheless, this work serves to fill the gap with respect to diverse PA implementation that exemplifies what works for PST in various contexts, which is a novel contribution to decision-making in teacher education.

Conclusions

The main contribution of this review is to illustrate the current trends in PA of teaching performance in teacher education. The studies presented here showed that PA has been applied in all the years of teacher education with small and large groups, with summative and formative objectives, face-to-face and online. The assessment criteria of the studies mentioned differed in their nature and design, with most of the studies reporting benefits of PA. Nevertheless, in some cases, the study design could only associate outcomes with PA, rather than showing plausible evidence of a causative link. Thus, these results do not form conclusive evidence. This review suggests that significant effects of PA, found within experimental studies which incorporate PST as a specific application have not been widely demonstrated yet. However, the few studies with a measurable impact of PA on teaching performance found a notable increase. This gives stakeholders ideas about the expected results if well-designed PA is implemented.

In summary, student teachers need to learn how to teach and assess performance of peers during their professional life. PA provides an assessment and learning scenario to critically reflect and judge teaching performance, which might give them the tools to monitor their practice as well as peers', based on the internalization of criteria for teaching. In addition, the extent to which PA impacts teaching practices needs further empirical support.

Acknowledgements

Centro de Investigación para la Gestión Integrada del Riesgo de Desastres (CIGIDEN), ANID/ FONDAP/15110017.

Conflict of interests

The authors declare no conflict of interest.

References

Abdulwahed, S. (2011). Student teachers' microteaching experiences in a pre-service English teacher education program. *Journal of Language Teaching and Research, 2*(5), 1043-1051. https://doi.org/10.4304/jltr.2.5.1043-1051

Al-Barakat, A., & Al-Hassan, O. (2009). Peer assessment as a learning tool for enhancing student teachers' preparation. *Asia-Pacific Journal of Teacher Education, 37*(4), 399-413. https://doi.org/10.1080/13598660903247676

Amobi, F. A., & Irwin, L. (2009). Implementing on-campus microteaching to elicit pre-service teachers' reflection on teaching actions: Fresh perspective on an established practice. *Journal of the Scholarship of Teaching and Learning, 9*(1), 27-34. Retrieved from: https://scholarworks.iu.edu/journals/index.php/josotl/article/view/1712

Arnold, L., Shue, C. K., Kalishman, S., Prislin, M., Pohl, C., Pohl, H., & Stern, D. T. (2007). Can there be a single system for peer assessment of professionalism among medical students? A multi-institutional study. *Academic Medicine, 82*(6), 578-586. https://doi.org/10.1097/ACM.0b013e3180555d4e

Arnold, L., Shue, C. K., Kritt, B., Ginsburg, S., & Stern, D. T. (2005). Medical students' views on peer assessment of professionalism. *Journal of General Internal Medicine, 20*(9), 819-824. https://doi.org/10.1111/j.1525-1497.2005.0162.x

Asghar, A. (2009). Reciprocal peer coaching and its use as a formative assessment strategy for first-year students. *Assessment & Evaluation in Higher Education, 35*(4), 403-417. https://doi.org/10.1080/02602930902862834

Britton, L. R., & Anderson, K. A. (2010). Peer coaching and pre-service teachers: Examining an underutilized concept. *Teaching and Teacher Education, 26*(2), 306–314. https://doi.org/10.1016/j.tate.2009.03.008

Bunch, G. C., Aguirre, J. M., & Tellez, K. (2009). Beyond the scores: Using candidate responses on high stakes performance assessment to inform teacher preparation for English learners. *Issues in Teacher Education, 18*(1),103-128. Retrieved from: https://files.eric.ed.gov/fulltext/EJ851544.pdf

Cabello, V. M. & Topping, K. J. (2018) Making scientific concepts explicit through explanations: Simulations of a high-leverage practice in teacher education. *International Journal of Cognitive Research in Science, Engineering and Education, 6*(3), 35-47 https://doi.org/10.5937/ijcrsee1803035C

Cabello, V. M. (2017). Role-playing for learning to explain scientific concepts in teacher education. Journal of Science Education, 18(2). 67-70 Retrieved from: http://chinakxjy.com/downloads/V18-2017-2/V18-2017-2-6.pdf

Charalambous, C., Hill, H., & Ball, D. (2011). Prospective teachers' learning to provide instructional explanations: how does it look and what might it take? *Journal of Mathematics Teacher Education, 14*(6), 441-463. https://doi.org/10.1007/s10857-011-9182-z

Cook, D. A., & West, C. P. (2012). Conducting systematic reviews in medical education: a stepwise approach. *Medical Education, 46*(10), 943-952. http://doi.org/10.1111/j.1365-2923.2012.04328.x

d'Alessio, M. A. (2018). The Effect of Microteaching on Science Teaching Self-Efficacy Beliefs in Preservice Elementary

Teachers. *Journal of Science Teacher Education, 29*(6), 441-467. https://doi.org/10.1080/1046560X.2018.1456883
Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. J*ournal of Teacher Education, 57*(2), 120-138. https://doi.org/10.1177/0022487105283796
Day, C. (2008). Committed for life? Variations in teachers' work, lives and effectiveness. *Journal of Educational Change, 9*(3), 243-260. https://doi.org/10.1007/s10833-007-9054-6
Duran, D., & Monereo, C. (2008). The Impact of Peer Tutoring on the Improvement of Linguistic Competence, Self-Concept as a Writer and Pedagogical Satisfaction. *School Psychology International, 29*, 481-499. https://doi.org/10.1177/0143034308096437
Eather, N., Riley, N., Miller, D., & Jones, B. (2017). Evaluating the effectiveness of using peer-dialogue assessment (PDA) for improving pre-service teachers' perceived confidence and competence to teach physical education. *Australian Journal of Teacher Education, 42*(1), 69-83. https://doi.org/10.14221/ajte.2017v42n1.5
Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. London: Routledge Falmer.
Feiman-Nemser, S. (2008). Teacher learning: How do teachers learn to teach? In M. Cochran-Smith, S. Feiman-Nemser, D. McIntyre & K. Demers (Eds.), *Handbook of research on teacher education. Enduring questions in changing contexts* (pp. 697-705). New York: Routledge.
Gielen, S., Dochy, F., & Onghena, P. (2010). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*(2), 137-155. https://doi.org/10.1080/02602930903221444
Gielen, S., Dochy, F., Onghena, P., Struyven, K., & Smeets, S. (2011). Goals of peer assessment and their associated quality concepts. *Studies in Higher Education, 36*(6), 719-735. https://doi.org/10.1080/03075071003759037
Harford, J., & MacRuairc, G. (2008). Engaging student teachers in meaningful reflective practice. T*eaching and Teacher Education, 24*(7), 1884-1892. https://doi.org/10.1016/j.tate.2008.02.010
Hume, A. C. (2012). Primary Connections: Simulating the classroom in initial teacher education. *Research in Science Education, 42*, 551-565. https://doi.org/10.1007/s11165-011-9210-0
Inoue, N. (2009). Rehearsing to teach: content-specific deconstruction of instructional explanations in pre-service teacher training. *Journal of Education for Teaching, 35*(1), 47-60. https://doi.org/10.1080/02607470802587137
Jian, W., Odell, S. J., & Schwille, S. A. (2008). Effects of teacher induction on beginning teachers' teaching. *Journal of Teacher Education, 59*(2), 132-152. https://doi.org/10.1177/0022487107314002
Kiliç, D. (2016). An Examination of Using Self-, Peer-, and Teacher-Assessment in Higher Education: A Case Study in Teacher Education. *Higher Education Studies, 6*(1), 136-144. Retrieved from: https://eric.ed.gov/?id=EJ1099387
Kilic, G. B., & Cakan, M. (2007). Peer assessment of elementary science teaching skills. *Journal of Science Teacher Education, 18*(1), 91-107. https://doi.org/10.1007/s10972-006-9021-8
Kim, M. (2009). The impact of an elaborated assessee's role in peer assessment. *Assessment & Evaluation in Higher Education, 34*(1), 105-114. https://doi.org/10.1080/02602930801955960
Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*(4), 344-348. https://doi.org/10.1016/j.learninstruc.2009.08.005
Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(2), 193-211. https://doi.org/10.1080/02602938.2019.1620679
Lin, G.-Y. (2018). Anonymous versus identified peer assessment via a Facebook-based learning application: Effects on quality of peer feedback, perceived learning, perceived fairness, and attitude toward the system. *Computers & Education, 116,* 81-92. https://doi.org/10.1016/j.compedu.2017.08.010
MacArthur, C. A., Schwartz, S. S., & Graham, S. (1991). Effects of a reciprocal peer revision strategy in special education classrooms. Learning Disabilities Research and Practice, 6, 201-210. https://psycnet.apa.org/record/1992-29343-001
McNulty, M. (2019). Peer Teaching Does Not Influence Performance in an Interprofessional Anatomy Course. *FASEB Journal, 33*(1). https://doi.org/10.1096/fasebj.2019.33.1_supplement.328.4
Mercader, C., Ion, G., & Díaz-Vicario, A. (2020). Factors influencing students' peer feedback uptake: instructional design matters. *Assessment & Evaluation in Higher Education*, 1-12. https://doi.org/10.1080/02602938.2020.1726283
Norcini, J. J. (2003). Peer assessment of competence. *Medical Education, 37*(6), 539-543. https://doi.org/10.1046/j.1365-2923.2003.01536.x
Okhremtchouk, I., Seiki, S., Gilliland, B., Atch, C., Wallace, M., & Kato, A. (2009). Voices of pre-service teachers: Perspectives on the Performance Assessment for California Teachers (PACT). *Issues in Teacher Education, 18*(1), 39-62. Retrieved from: https://files.eric.ed.gov/fulltext/EJ851541.pdf
Oluwatayo, J. A., & Adebule, S. O. (2012). Assessment of teaching performance of student-teachers on teaching practice. *International Education Studies, 5*(5), 109-115. Retrieved from: https://files.eric.ed.gov/fulltext/EJ1067071.pdf
Oren, F. S. (2012). The effects of gender and previous experience on the approach of self and peer assessment: A case from Turkey. *Innovations in Education and Teaching International, 49*(2), 123-133. https://doi.org/10.1080/14703297.2012.677598
Ostrosky, M. M., Mouzourou, C., Danner, N., & Zaghlawan, H. Y. (2013). Improving teacher practices using microteaching: Planful video recording and constructive feedback. *Young Exceptional Children, 16*(1), 16-29. https://doi.org/10.1177/1096250612459186
Pecheone, R. L., & Chung, R. R. (2006). Evidence in teacher education: The Performance Assessment for California Teachers (PACT). *Journal of Teacher Education, 57*(1), 22-36. https://doi.org/10.1177/0022487105284045
Ratminingsih, N. M., Artini, L. P., & Padmadewi, N. N. (2017). Incorporating Self and Peer Assessment in Reflective Teaching Practices. *International Journal of Instruction, 10*(4), 165-184. https://doi.org/10.1177/0022487105284045
Sandholtz, J. H., & Shea, L. M. (2012). Predicting Performance: A Comparison of University Supervisors' Predictions and Teacher Candidates' Scores on a Teaching Performance Assessment. *Journal of Teacher Education, 63*(1), 39-50. https://doi.org/10.1177/0022487111421175
Seifert, T., & Feliks, O. (2019). Online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills. *Assessment & Evaluation in Higher Education, 44*(2), 169-185. https://doi.org/10.1080/02602938.2018.1487023
Sluijsmans, D., & Prins, F. (2006). A conceptual framework for integrating peer assessment in teacher education. *Studies in*

Educational Evaluation, 32(1), 6-22. https://doi.org/10.1016/j.stueduc.2006.01.005
Sluijsmans, D., Brand-Gruwel, S., van Merriënboer, J. J. G., & Bastiaens, T. J. (2003). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation, 29*(1), 23-42. Retrieved from: https://eric.ed.gov/?id=EJ670665
Sluijsmans, D., Brand-Gruwel, S., van Merriënboer, J. J. G., & Martens, R. L. (2004). Training teachers in peer-assessment skills: Effects on performance and perceptions. *Innovations in Education and Teaching International, 41*(1), 59-78. https://doi.org/10.1080/1470329032000172720
Sonmez, D., & Can, M. H. (2010). Preservice science teachers' ability to identify good teaching practices. *Procedia - Social and Behavioral Sciences, 2*(2), 4120-4124. https://doi.org/10.1016/j.sbspro.2010.03.650
Stiggins, R. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice, 10*, 7-12. https://doi.org/10.1111/j.1745-3992.1991.tb00171.x
Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*(4), 265-269. https://doi.org/10.1016/j.learninstruc.2009.08.002
Sun, Q., Wu, J., Rong, W., & Liu, W. (2019). Formative assessment of programming language learning based on peer code review: Implementation and experience report. *Tsinghua Science and Technology, 24*(4), 423-434. https://doi.org/10.26599/TST.2018.9010109
Topping, K. J. (2009). Peer assessment. *Theory Into Practice, 48*(1), 20-27. https://doi.org/10.1080/00405840802577569
Topping, K. J. (2010). Peers as a source of formative assessment. In H. Andrade / G. J. Cizek (Eds.), Handbook of formative assessment (pp. 61-74). New York: Routledge. Retrieved from: https://eric.ed.gov/?id=ED579876
Tsai, C., Lin, S. S. J., & Yuan, S.-M. (2002). Developing science activities through a networked peer assessment system. *Computers & Education, 38*(1–3), 241-252. https://doi.org/10.1016/S0360-1315(01)00069-0
van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*(1), 41-54. https://doi.org/10.1016/j.edurev.2008.11.002
van Zundert, M., Sluijsmans, D., & van Merriënboer, J. J. G. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270-279. https://doi.org/10.1016/j.learninstruc.2009.08.004
Vermunt, J., & Endedijk, M. (2011). Patterns in teacher learning in different phases of the professional career. *Learning and Individual Differences, 21*, 294-302. https://doi.org/10.1016/j.lindif.2010.11.019
Vickerman, P. (2009). Student perspectives on formative peer assessment: An attempt to deepen learning? *Assessment & Evaluation in Higher Education, 34*(2), 221-230. https://doi.org/10.1080/02602930801955986
Wen, M. L., & Tsai, C. (2008). Online peer assessment in an in-service science and mathematics teacher education course. *Teaching in Higher Education, 13*(1), 55-67. https://doi.org/10.1080/13562510701794050