



Investigating peer-assessment strategies for mathematics pre-service teacher learning on formative assessment

Michal Ayalon¹ · Karina J. Wilkie²

Published online: 27 May 2020
© Springer Nature B.V. 2020

Abstract

Formative assessment practices for secondary mathematics have been advocated as valuable for students, but difficult for teachers to learn. There have been calls in the literature to increase the emphasis on formative assessment in mathematics teacher preparation courses. This study explored the use of peer-assessment strategies for helping pre-service secondary mathematics teachers (PSTs) cultivate formative assessment principles and practices for assessing school students. Twenty-seven PSTs participated in a peer-assessment cycle comprised of: sourcing a rich mathematics task; constructing an assessment rubric for it; and collecting and analysing a selection of secondary student responses to the task. Each PST then provided written and verbal feedback to a peer on his/her rubric and student solution assessments. We draw on theoretical conceptions of Teacher Assessment Literacy in Practice to characterize the PSTs' perceptions of their experience of formative assessment processes for learning to assess school students, in terms of cognitive and affective dimensions of their conceptions of assessment. The cohort evidenced a wide range of levels of confidence with the various aspects of formative assessment practices but on average less confidence in assessing school student task responses themselves than in assessing peer work. In addition to highlighting specific changes to different types of assessment knowledge, the PSTs also evidenced an awareness of shifts in their attitudes, in coming to view student task responses with more appreciation and humility.

Keywords Formative assessment · Peer assessment · Pre-service teacher education · Rubric construction · Secondary mathematics

Formative assessment may be described as a pedagogical approach that emphasizes the importance of both teachers and students using students' understanding to inform instruction and promote learning (Black and Wiliam 2009). The primary purpose of formative

✉ Michal Ayalon
mayalon@edu.haifa.ac.il

Karina J. Wilkie
karina.wilkie@monash.edu

¹ Department of Mathematics Education, University of Haifa, Haifa, Israel

² Faculty of Education, Monash University, 47-49 Moorooduc Highway, Frankston, Melbourne, VIC 3199, Australia

assessment is not to assign scores or grades, but rather to highlight conceptual strengths and challenges faced by the students. Teaching that connects and builds upon students' mathematical thinking has been associated with classroom interactions and discourse that promote collaboration and conceptual understanding (e.g. Ball and Forzani 2009; Cobb et al. 1991) and improvement in student achievement (Fennema et al. 1996; Wiliam et al. 2004). Similarly, opportunities centred on teachers' understanding of their students' thinking to inform future teaching can lead to changes in teachers' pedagogical beliefs (Cobb et al. 1991; Fennema et al. 1996) and discourse about student thinking (Walkoe 2015).

The literature typically specifies several key elements of formative assessment (e.g. Black and Wiliam 2009; Heritage 2007; Sadler 1998). These include setting tasks and sharing learning intentions and criteria for success; eliciting evidence of student understanding and identifying the 'gap' between a student's current understanding and the desired goals; providing feedback that moves learners forward; and activating students as instructional resources for one another and as owners of their own learning. Implementation of formative assessment practices poses substantial challenges for teachers. Researchers and recent reforms in mathematics education have emphasized the need for practising teachers to develop professional expertise in assessing and using student thinking to guide instruction (Ball and Forzani 2009; Evans and Ayalon 2016; National Council of Teachers of Mathematics [NCTM] 2000). Research has shown that prior to hands-on experiences with classroom assessment, prospective teachers commonly hold views of assessment that are largely summative—assessment is equated with the administering tests and assigning grades (Graham 2005). Such conceptions often evidence a limited knowledge of formative assessment principles and practices (MacLellan 2004). Thus, there is a need for assessment training to begin as early as possible in teacher education (Heritage 2007; Shepard 2005).

Teacher conceptions of assessment have both cognitive and affective aspects, which are framed by teachers' epistemological beliefs as well as broader views of teaching and learning (Xu and Brown 2016). Xu and Brown (2016) drew on Stiggins' (1991) definition of Assessment Literacy (AL) to encompass a combination of cognitive dimensions, such as teachers' knowledge about assessment and what teachers believe is true and false about assessment (as was offered in Stiggins' definition), and affective aspects such as emotional inclinations that teachers have about various aspects and uses of assessment. Xu and Brown (2016) highlighted that 'To improve teacher AL [in both cognitive and affective terms] inevitably involves a long process of attending to, and possibly changing, teachers' existing conceptions of assessment' (Xu and Brown, p. 156). To promote development of assessment practices, there has been a recent movement in teacher education programs towards practice-based approaches that structure prospective teachers' experiences around critical tasks and problems that echo teachers' daily work (Authors, first online; Ghousseini and Herbst 2016). The current study takes this approach to further learn about possible ways to cultivate pre-service secondary mathematics teachers' assessment literacy with a focus on formative assessment practices.

In this study, a group of PSTs participated in a cycle of peer-assessment activities where they experienced two roles: (1) assessors of actual secondary students' solutions to selected rich mathematical tasks and (2) feedback providers to their peers (other PSTs) on their constructed rubrics and student assessments. This cycle's design drew on several aspects, identified in previous studies as effective for teachers' professional learning, such as participants developing assessment tasks and rubrics, analysing student work, and teamwork. The uniqueness of this particular approach is in approximating teachers' actual practice in schools (Grossman et al. 2009) by using peer-assessment processes designed specifically to

give PSTs experience with authentic formative assessment practices, but at a reduced level of complexity suitable for pre-service education.

Much of the literature on peer assessment as a learning strategy in teacher education (e.g. Sluijsmans et al. 2004; Zevenbergen 2001) context has focused on PSTs' development of Content Knowledge (e.g. mathematics problem solving). This study aimed to explore the potential of using peer assessment with pre-service mathematics teachers for learning Pedagogical Content Knowledge (PCK) (Shulman 1986), specifically formative assessment principles and practices. Peer assessment is envisaged as approximating for PSTs (Grossman et al. 2009) the collaborative teamwork of practising teachers in school contexts. We seek to understand more about prospective teachers' experiences and perceptions of such an approach in both terms of cognitive and affective dimensions as a way to gain initial insights into the implementation of practice-based learning for developing formative assessment practices in secondary mathematics. In attending to the perspectives of the prospective teachers who took part in the peer-assessment cycle experience, this exploratory study addresses the following research question: *How do mathematics prospective teachers perceive their experience of formative peer-assessment processes for learning to assess school students, in terms of cognitive and affective dimensions of their conceptions of assessment?*

Theoretical perspectives

In developing this study, we drew on theoretical perspectives in the literature on the types of formative assessment knowledge needed by teachers, conceptualizations of teachers' assessment literacy development, and processes for pre-service teacher professional learning. These are overviewed in turn in the following subsections.

Definition of formative assessment and key teacher practices

There is widespread agreement in the literature about the importance of formative assessment for effective learning, but still-evolving definitions and differing views on its operationalisation in various learning contexts and disciplines (Black 2015; Popham 2009; Taras 2007). The definition we chose for this study views assessment as formative 'to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited' (Black and Wiliam 2009, p. 6).

Formative assessment that seeks to understand students' current achievement for the purposes of giving effective feedback and informing future teaching is conceptualized in Wiliam and Thompson's (2007) proposed framework of five key teacher practices:

1. Clarifying, sharing, and understanding learning intentions and criteria for success;
2. Engineering effective classroom discussions, questions, and tasks that elicit evidence of learning;
3. Providing feedback that moves learners forward;
4. Activating students as instructional resources for one another; and
5. Activating students as the owners of their own learning.

Formative assessment relies on teachers being able to provide constructive task-specific feedback for students to follow up on to improve their learning (Sadler 1998). Studies have shown that students often find assessment criteria for a rich task difficult to interpret (Black and Wiliam 2009). One strategy in the literature is teachers providing a rubric to students that highlights the specific learning intentions for a task and what it means to achieve the highest level of success (Black et al. 2003; Schafer et al. 2001). Although in this study we did not include PSTs actually giving feedback to students with their rubrics or using it to inform future teaching, we considered the process of constructing rubrics and assessing real student responses to be an important precursor to the PSTs’ future in-service development.

Conceptualizations of teachers’ assessment literacy development

It has been argued that there is a need for research on teachers’ assessment literacy to inform good assessment education for prospective teachers (DeLuca and Klinger 2010). An early definition of assessment literacy (AL) related to applying a basic understanding of assessment to various measures of student achievement, to being able to distinguish between low- and high-quality assessments, and using those that communicate clear, specific, and rich definitions of the intended achievement (Stiggins 1991). More recently, the concept of assessment literacy has been expanded to encompass pre-service teacher development. Xu and Brown (2016) synthesized 100 studies of AL from both educational assessment and teacher education fields of research to develop the triangular model presented in Fig. 1.

The TALiP model emphasizes the processes of co-construction, participation, and reflection by which teachers (pre- and in-service) develop their knowledge base over time and experience both cognitive and affective adjustments to their views of learning and conceptions of assessment. The foundational types of assessment knowledge listed along the base of Xu and Brown’s (2016) triangular TALiP model can be applied to the previously presented framework of formative teacher practices (Wiliam and Thompson 2007). For example, selection of a high-quality assessment task, which has clear criteria for achievement linked to its learning intentions (strategy 1) and will elicit evidence of student learning (strategy 2), requires *Disciplinary knowledge and PCK* as well as *Knowledge of assessment purposes, content, and methods*. Interpretation of student responses and

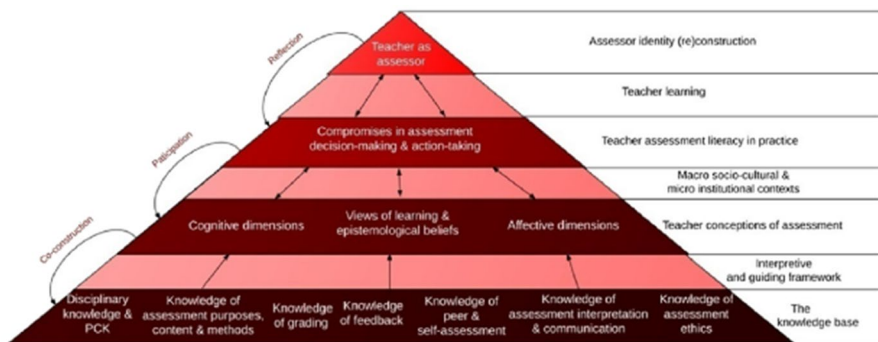


Fig. 1 Model on developing Teacher Assessment Literacy in Practice (TALiP). (Source: Xu and Brown 2016, p. 155)

communicating effective feedback (strategy 3), for example addressing a misconception or extending their knowledge of a particular concept, additionally requires *Knowledge of assessment interpretation and communication*, *Knowledge of grading*, and *Knowledge of feedback*. To activate students as instructional resources for each other (strategy 4) and as owners of their own learning (strategy 5) in ways that are socially just, equitable, and inclusive, teachers also need *Knowledge of peer and self-assessment* and *Knowledge of assessment ethics*.

In being educated about assessment, teachers mediate their uptake of knowledge by filtering and interpreting it with their ‘conceptions of assessment’ (third bottom layer of the triangular model in Fig. 1): a guiding framework that encompasses both cognitive and affective responses. The cognitive dimension includes what teachers believe is true and false about assessment. Teachers tend to adopt new knowledge, ideas, and strategies for assessment that are congruent with their conceptions of assessment, while rejecting those that are not (Xu and Brown 2016). The affective dimension signifies emotional inclinations that teachers have about various aspects and uses of assessment. Teachers may have strong and weak, positive and negative emotions about assessment that arise from various experiences over time (Crossman 2007). Teacher conceptions are considered both individualized and collective, due to the social nature of education. The emotional dimension of conceptions may make conceptual change difficult, leading to less effective learning about assessment and reduced effectiveness in implementing new assessment policies (Green 1971).

The apex of the triangular TALiP model denotes the ultimate goal of developing teachers’ assessment literacy: that the teacher’s role as ‘assessor’ is successfully integrated into their professional *identity*. Xu and Brown (2016) argued that this identity development process requires a new way of understanding what it means to be a teacher. Looney et al. (2018) offered the term ‘Teacher Assessment Identity’ (TAI) to conceptualize teachers’ identity development in their role of assessor. Their TAI framework includes five interconnected aspects, with each expressed in the first person to emphasize their self-reflective nature: *I know* and *My role* (which in this study we considered as cognitive), and *I believe*, *I am confident*, and *I feel* (which we considered affective). Researchers disagree about conceptualizations of belief: to align them with the affective domain or with cognition (Philipp 2007). Yet there has been consensus that in educational research, differentiating between teacher knowledge and beliefs is important (Philipp 2007). For the purposes of our study, we have drawn on McLeod’s (1992) definition of the affective domain as ‘a wide range of beliefs, feelings, and moods that are generally regarded as going beyond the domain of cognition’ (p. 576). Unlike traditional definitions of assessment literacy that focus on the cognitive domain (e.g. Stiggins 1991), Looney et al.’s (2018) model adds the emotional and dispositional dimensions of assessment identity development, similar to Xu and Brown’s (2016) inclusion of both cognitive and affective dimensions in their layer ‘Teacher conceptions of assessment’. In our study, we sought evidence of the PSTs expressing cognitive and affective dimensions of their experiences of formative assessment practices in ways related to these five aspects.

Pre-service teacher learning through approximations of practice

In the last decade, the literature on teacher education has argued for preparation programs that better support teachers to develop and deploy knowledge and skills relevant to their actual practice (Chick and Beswick 2017; Grossman et al. 2009; Popham 2009). A study of cross-professional courses led Grossman et al. (2009) to conclude that the field of teacher

education needed to develop *pedagogies of enactment*, joining them with existing pedagogies of investigation (e.g. Merseth 1991). The fundamental aim of such a proposal is to better support teachers in learning how to use knowledge in action (Ball and Forzani 2009). In recent years, a growing group of scholars in the field have been experimenting with strategies that organize the work and scholarship of teacher education around what they refer to as *core practices* of classroom teaching. By highlighting specific, routine aspects of teaching that demand the exercise of professional judgment and the creation of meaningful intellectual and social community for teachers, teacher educators, and students, these core practices may offer pre-service teacher educators powerful tools for preparing teachers for the constant in-the-moment decision-making that the profession requires (Grossman et al. 2009). Grossman et al. (2009) conceptualized teacher preparation courses as involving PSTs in *approximations of practice*, which are authentic but discrete components of teaching practice with reduced complexity. In this study, we sought to include approximations of formative assessment practices (William and Thompson 2007), such as selecting a rich task for eliciting evidence of learning, developing criteria for success, and assessing (real) student responses. We also conceptualized peer assessment as an approximation of practice and sought to explore its potential as a learning tool for developing PSTs' assessment literacy. This is described in the following subsection.

Proposing peer assessment as an approximation of practice

Peer assessment (PA) is an educational arrangement where learners consider and specify the level, value, or quality of a product or performance of other equal-status learners (Topping and Ehly 1998). Numerous studies, particularly in higher education settings, have tended to focus on the reliability and validity of peer (undergraduate student) assessment for summative grading of course assignments, compared to lecturer assessment. Yet there is an increasing interest in peer assessment as a complementary component of *formative* assessment practices (Black and William 2009; Kollar and Fischer 2010; Topping 2010). Some studies researched qualitative peer feedback, and they found evidence of its potential for effective learning. Peer-assisted learning can be more immediate, timely, and individualized than teacher feedback (Topping 2010). It has also been found to develop self-regulation and metacognition, improve communication skills, yield learners' better understanding of the assessment criteria, and increase self-awareness of the quality of one's own work (Andrade and Cizek 2010; Brown and Harris 2013; Sadler 1998; Topping 2010). When learners analyse the work of other learners, they have access to a variety of examples that help them better see gradations in quality (Topping 1998). Moreover, because they have not created the work themselves, they view it from a more distanced perspective and seem able to analyse it more objectively, compared to self-assessment (Black et al. 2003). It has been found that peer feedback can also be beneficial to learning because it is qualitatively different from the usual teacher feedback: the absence of a clear 'knowledge authority' and the uncertainty induced by a peer's relatively equal status may induce learners to search for confirmation of the feedback, leading to further thinking and discussion (Yang et al. 2006).

We found a few studies (Sluijsmans et al. 2004; Zevenbergen 2001) that investigated the potential of using peer assessment as a learning tool specifically with prospective teachers. These studies aimed to research improvements to PSTs' subject matter knowledge per se, rather than on their assessment literacy development. Yet their findings also suggested that in addition to improving PSTs' knowledge, peer-assessment provided valuable experience

in learning to make and communicate qualitative judgements: an important teacher assessment practice.

Topping (2010) included the affective dimension in his categories of proposed influences on the effectiveness of peer-assessment processes. He primarily described it in terms of social interactions between peers, establishing a trusting relationship, and motivational issues. Sluijsmans et al. (2004) study of PSTs included Likert-scale response items on perceptions of 'group atmosphere' and 'group behaviour'. Yet to date there appears to be very little in the literature that researches peer assessment for developing prospective teachers' assessment literacy or affective aspects of their peer-assisted learning on such development.

In this study, we conceptualize peer-assessment processes as an authentic approximation of teachers' formative assessment practice—the cycles of collaboration and critique necessary in schools for the core practice of teachers creating, implementing, and interpreting assessment tasks, particularly rich tasks for giving formative feedback to school students. And we explicitly attend to both cognitive and affective aspects related to their assessment literacy development. In exploring the potential for such processes for developing prospective teachers' assessment literacy, we adapted Reinholz's (2016) six-phase peer-assessment cyclical model, shared in the next section.

Research design

This section presents details about the peer-assessment cycle design used in the study, the participants, and the context for the study. We then describe the data collection and data analysis processes.

Peer-assessment cycle design

Reinholz's (2016) peer-assessment cycle model was originally developed in an undergraduate context to support students' learning and centres around six phases: (1) task engagement, (2) peer analysis, (3) feedback provision, (4) feedback reception, (5) peer conferencing, and (6) revision. In this study, we applied the model to PSTs' assessment literacy development and added reflection as a seventh and final phase, to incorporate self-reflection opportunities to help teachers understand the links between what they do and how they might improve their effectiveness (Looney et al. 2018; Xu and Brown 2016) (see Fig. 2).

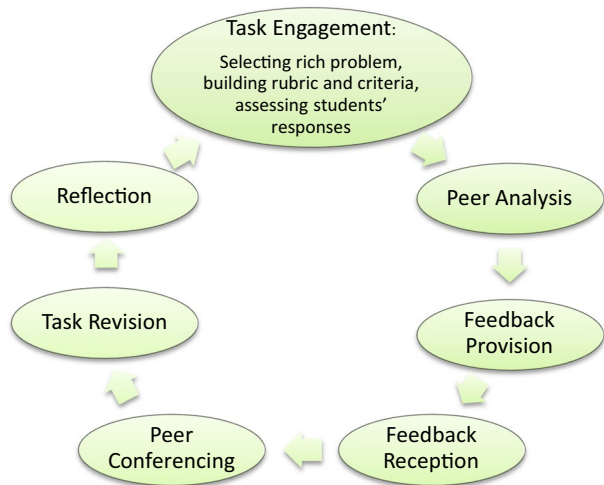
Task engagement

The cycle began with the PSTs experiencing several steps associated with teacher assessment practice: selecting a rich mathematics problem as an assessment task for students, attempting the problem themselves first, developing a rubric with criteria for assessment, and using the rubric for assessing real school student solutions to the problem (to identify all important aspects to assess and possible performance levels as advised by Danielson and Marquez (2016)).

Peer analysis

The peer analysis phase involved making judgements about the quality of the work of a peer (their designed rubric for assessing collected student solutions), to generate elaborated

Fig. 2 Study's peer-assessment cycle. (Adapted from Reinholz 2016)



and qualitative feedback, and give constructive criticism. This analytic experience was to support individuals in developing a sense of distanced objectivity for application to their own work (Black et al. 2003; Reinholz 2016). We reason that a PST who provides feedback to a peer on the criteria used to assess student solutions may begin to notice the importance of alignment among the purposes of assessment, the chosen mathematical task, and the assessment criteria. Because she is ‘a step removed’ from the particular work she is analysing, she is more likely to notice any discrepancies (van Gennip et al. 2010). Such observation can become a lens that she can later apply to her own work of assessment.

In analysing peer work, PSTs may also be exposed to a variety of examples, which helps them see variations in quality (Sadler 1998). This contrasts with presenting them with only high-quality assessment samples, which may make it difficult for them to determine what it actually is that makes an assessment ‘good’. Such experience, even with hypothetical work, may help the PSTs develop a deeper understanding of assessment.

Feedback provision

Feedback provision involved the PSTs in communicating feedback by writing their analyses for their peers (see "Appendix 2" for the instructions given for writing the feedback).

Feedback reception

Receiving feedback encourages an individual to view their work from another’s perspective, and to attend to aspects of their work that seem to be sound or problematic. Not all feedback is equally useful (Hattie and Timperley 2007). Feedback that helps participants learn to analyse, critique, and improve their own work independently is arguably of most benefit (Hattie and Timperley 2007). Simply receiving feedback that tells the correct answer with no explanation is likely to be of less value (Hattie and Timperley 2007). Even worse, feedback in the form of only a grade or praise that focuses individuals on themselves rather than the task has been associated with reduced motivation and performance (Butler 1988).

Peer conferencing

Peer conferencing was also included to encourage PSTs to discuss their feedback and analyses with each other. Through conversations, the participants could explain their ideas verbally and also discuss issues more broadly. Because the PSTs in each pair had already spent time thinking and working on the assessment activity before their conversation, even a short discussion was beneficial and productive. Conferencing also was intended to allow participants to provide affective support, compared to written feedback, which in being less nuanced, may be perceived as more critical or insensitive than intended (Patton 2012). Each PST was involved in two conferences, but each pair did not mutually assess each other's work: PST A assessed PST B's work, PST B assessed PST C's work, and so on. The aim of not working in a 'fixed' couple was to allow the participants to be exposed to more than one voice.

Task revision

After receiving the peer feedback, the participants had an opportunity to revise their work before submitting their finished product. When participants know that they will be expected to revise their work, it can influence the feedback they give as well as their perception of the feedback they themselves receive (Reinholz 2016).

Reflection

As a final phase in the peer-assessment cycle, the PSTs were invited to complete an individual reflective questionnaire. The questionnaire focused on both affective and cognitive dimensions of the participants' conceptions of assessment:

1. The PSTs were asked to score their perceived level of confidence (on scale of 1–5; 5 being the highest) for assessing school students' task solutions, and for providing feedback to peers, and to explain in detail their choices in terms of their individual strengths and challenges.
2. The PSTs were asked to write about any learning they perceived as gained from their participation in the peer-assessment cycle and to provide specific illustrations.

For additional insights related to the collective nature of teacher conceptions, a group interview was also conducted, focusing on the PSTs' experiences in the activities, their perceived strengths and difficulties, and their learning gains.

Participants and context for the study

A group of 27 prospective teachers participated in this study, which was conducted in Israel within their teacher preparation course on Theory and Practice in Teaching Mathematics: Algebra and Functions. Some of the PSTs were studying for a dual degree in Mathematics and Education, and several had already graduated with a B.Sc. from the university's Department of Mathematics. During the two sessions before the data collection for the study took place, the PSTs were introduced to the literature on definitions, theoretical ideas, and practices associated with formative assessment. These included the previously presented formative assessment definitions and characteristics, as well as various

mathematics-specific frameworks for assessing students' competency (e.g. Swan and Burkhart 2012; Schwartz et al. 1995). These frameworks (see "Appendix 1") reflect the values of the mathematics reform movement (NCTM 2000) and focus on rich, mathematically complex work that requires students to create plans, make decisions or solve problems, and then justify their thinking. The PSTs were provided with a selection of rich mathematics tasks and engaged in writing solutions and analysing them to define the learning goals and assessment criteria with rubrics. A rich task was defined for the PSTs as an open-response task that invites multiple solution strategies, reasoning, justification, and explanation. They discussed in small groups their ideas about possible levels of solution quality and expected student difficulties. As a class, they shared their proposed rubrics for each task and discussed them in terms of giving effective feedback to students on their current level of achievement and next steps in their learning. After these two introductory sessions, the PSTs participated in the peer-assessment cycle. They were informed that the assessments made by their peers were for their professional learning and would not be used at all in determining their summative grades for that course.

Data collection and analysis

The peer-assessment cycle lasted for two sessions (total of 5 h). The first session, lasting for 2 h, focused on the first phase of the cycle: engagement with the task. The PSTs were each asked to select a rich mathematics problem they wanted to use as an assessment task for school students, decide on its assessment purpose/s, and to construct a rubric with appropriate assessment criteria. They were then asked to give the mathematics problem to five or six school students, and to assess each response using their rubric. For the next session (after a week), the PSTs were asked to bring a report detailing: the chosen mathematics problem, the constructed rubric, the school student solutions, and their assessment of each solution. They were informed beforehand that they would be engaged in a peer-assessment activity and encouraged to write their reports clearly and articulately. The written work products from these sessions were included in the data collection.

The second session, lasting for 3 h, focused on the subsequent phases of the peer-assessment cycle. First of all, each PST received one peer's report and was asked to read it carefully and provide written feedback. This was followed by each participant receiving written feedback on their report and spending time reading and reflecting on it. Then, each pair of PSTs (assessor and assessed) met for a conference. As mentioned, this activity occurred twice, with each participant engaging in one conference as the assessor, and in another conference as the assessed. After discussing the assessments and feedback, the PSTs were asked to return to their initial report and make revisions to improve their work, and these were also collected for comparison with the original reports as evidence of their knowledge development, and to triangulate with the PSTs' own perceptions of their learning, elicited with a written task. The PSTs were asked to reflect in detail on their experiences enacting two different roles: first as assessors of school students' responses to their mathematics problem and second as providers of feedback to their peers on their rubrics and student assessments (see stage 7 in the above-described peer-assessment cycle). In order to learn about the PSTs' perceptions of these formative peer-assessment processes for learning to assess school students, and in terms of cognitive and affective dimensions, the data collected included their perceived levels of confidence in assessing school students' mathematics responses and peers' work on a 5-point scale and their reflective responses to various open-ended prompts.

Finally, there was a whole-class conversation about the entire process. The PSTs discussed their learning about the various assessment criteria to include in a rubric for a rich mathematics task, such as: *understanding of the problem* (e.g. instructions, situation given, data and constraints), *mathematical thinking* (e.g. generalization, justification), *construction of a model*, *computations*, *manipulations*, *communication* (e.g. explanations, presentation of the solution, mathematical language), *creativity* (multiple solutions and originality), and *drawing conclusions*. These criteria were collectively posted and sorted on the board through discussion.

The aim of the data analysis was to explore the PSTs' perspectives on their experience of the peer-assessment cycle for learning to assess school students formatively with rich tasks, in terms of cognitive and affective dimensions of their conceptions of assessment. As a first step, we examined the PSTs' scores for levels of confidence. Then, we conducted interpretive and in-depth qualitative analysis on the PSTs' written reflective narratives (Creswell 2007). We drew on Looney et al.'s (2018) conceptualization of Teacher Assessment Identity (TAI) for analysing the different aspects the PSTs evidenced in their reflections and narratives, in terms of their expression of feelings [I feel], beliefs [I believe], and self-efficacy [I am/am not confident] (all of which we considered in this study as affective), and role enactment [My role] and knowledge [I know] (which we considered as cognitive). Within the 'I know' dimension, we also looked for references to the specific types of assessment knowledge from the foundation base Xu and Brown's (2016) Teacher Assessment Literacy (TALiP) model.

The analysis involved iterations of sorting the data, and continual comparisons between the data and the developing categories, as well as across the categories themselves. This process resulted in a coding scheme with ten themes; four relate to cognitive dimensions and six relate to affective dimensions, presented in "Appendix 3". It is important to note that a PSTs' written reflective narrative could sometime express both affective and cognitive aspects of conceptions of assessment. For example, when asked about the learning gains from her participation in the peer-assessment cycle, PST8 wrote:

Throughout the experience, I have learnt to select good rich tasks and to build appropriate criteria for assessment [I know]. However, something bothers me [I feel]. When discussing my assessment with my colleague, I found it very confusing [I feel] that one of us evaluated a solution as excellent, whereas the other evaluated it as a medium level of quality. These differences of opinions made me feel uncomfortable and nervous [I feel]. What will I do when I become a real teacher? How will I know that I have interpreted the student's response correctly? [I feel]. (PST8)

In her response, PST8 referred to both cognitive and affective aspects of perception of her experience. She wrote that participating in the peer-assessment cycle supported the development of her skills in being able to design rich tasks and assessment criteria. We coded this expression as *I know* associated with Theme C1: Increased knowledge about designing rubrics for rich tasks (see findings section). PST8 also expressed confusion and anxiety as for utilizing the assessment tools to determine the quality levels of solutions. We coded these expressions as *I feel* associated with Theme A4: Less confidence in grappling with the subjective nature of certain assessment criteria ((see findings section). Decisions about interpretation were made collaboratively between the authors and consensus reached through discussion of noticeable patterns across the PSTs' reflections and specific nuances in their choice of wording.

Taking these ten themes from analysis of the PSTs' reflections into account, the group interview was then analysed. All of the themes that had previously been drawn from the

written reflections were also identified in the group interview, and no new themes emerged. The PSTs' written assessments (both the initial assessments designed in the first phase of task engagement and the revised versions generated after the peer-assessment), written feedback to their peers, and the notes taken during the peer-conference phase were also examined in detail to better understand the PSTs' descriptions and explanations in their reflections. These are drawn on in the Findings section to contextualize the participants' articulations.

Findings

This section addresses the study's research question on prospective teachers' perceptions of their experience of learning to assess school students' responses to mathematics tasks through a formative peer-assessment process, in terms of cognitive and affective dimensions; each is the focus of a one subsection. The five dimensions from Looney et al.'s (2018) model for Teacher Assessment Identity have been included in the PSTs' quotes (in square brackets) to highlight the cognitive and affective dimensions of their reflections about their experiences with assessment practices.

Cognitive aspects of the PSTs' perceived experience

Overall, the PSTs' reflections on salient learning outcomes evidenced their perceptions that the formative peer-assessment processes used in the study did enhance their learning. Four cognitive-related themes (C1–C4; C stands for Cognitive) emerged from analysis of the PSTs' reflections and were evident in the group interview responses; they also seemed to be agreed on by the group as a whole.

C1. Increased knowledge about designing rubrics for rich tasks One key theme that that nearly three-quarters of the PSTs referred to in their written reflections (and in the group interview) was their perception that participating in the peer-assessment cycle supported the development of their skills in being able to design appropriate assessment criteria for assessing students' responses to rich tasks. This seemed to be related by the PSTs to their discussion and negotiation with peers about the rubrics and assessments, rather than simply being about having further experience with assessment practices per se. In particular, peer critique of their work seemed to draw their explicit attention to relationships between their assessment goals and the mathematical task at hand, through comments and suggestions for removing, adding, or changing the weighting of different criteria. The two criteria (previously identified in perceived difficulties)—communication and creativity—often received attention in the peers' feedback to each other, with other criteria appearing more sporadically (for example, manipulation, problem understanding, and generalization). It appeared that having the opportunity for a feedback conference enabled PSTs to wrestle collaboratively with their difficulties.

For example, PST15 had chosen a task on finding examples of positive and negative numbers whose substitution in the algebraic expression $6 - a$ would result in positive values or negative values. One goal set by PST15 was to assess students' ability to reach generalizations related to the expression's behaviour (e.g. substitutions of numbers larger than 6 will result in negative values). Another goal was to assess students' communication, in particular their justification for their solution. Accordingly, two of the assessment criteria she defined were 'mathematical thinking: generalization' and 'communication'. When

assessing the students' solutions, she ranked several as 'low' on these two criteria, explaining that the students had only provided examples of numbers for each requirement, but without any attempt to reach a generalization. PST2 provided peer feedback to PST15's assessment. In her feedback, she wrote that the assessment goals and criteria set by PST15 did not match the task's requirements. She suggested modifying the task to ask explicitly for generalization. She also suggested adding the criterion 'computation', and to reconsider the inclusion of 'communication' because she thought that the explanation would be evident enough in a given generalization and not require further explanation. When reading PST2's feedback, PST15 accepted her task modification suggestion. However, she rejected PST2's idea of adding the criterion of 'computation', because for her, developing a sense of the behaviour of the algebraic expressions was most important. During their peer conference, they discussed the role of explanation in this task. They reached consensus on valuing a student's explanation, but not deducting points if correct generalizations are given.

When asked about the learning gains from her participation in the peer-assessment cycle, PST15 emphasized her 'new' knowledge:

The collaborative work allowed me to see what I had missed in the task and how I could change it so as to improve the assessment [I know]. It also opened my eyes to the complexity related to the expectation of explanations. Although it seems like a valued part of a solution, it is not clear whether all solutions to a task require it [I know]. (PST15)

This example highlights both the challenges inherent in rich task assessment and the benefits of peer assessment for approximating the negotiation processes teacher teams experience when selecting and moderating common assessment tasks.

C2. Knowledge of how to distinguish between levels of quality in responses Another theme evidenced in two-thirds of the PSTs' reflections as well as in the group interview was the contribution of participating in the peer-assessment cycle to improving their skills and knowledge for assessing tasks with rubrics. In particular, the peer feedback they received included explained approval of or disagreement with their judgments of students' solutions and suggestions for refinement, helping them to better differentiate the quality of different solutions. For example, PST1 chose a task on comparing the capacity of cylinders rolled ('landscape' or 'portrait') from the same-size (A4) paper (see Fig. 3).

One of the solutions given by a ninth-grade student was:

I will cut the A4 rectangle paper so as to make a square. Then I have two pieces of paper. One is a square and no matter how I roll it, it will have the same capacity. The other is another rectangle, and I will cut it again so as to make a square... then again

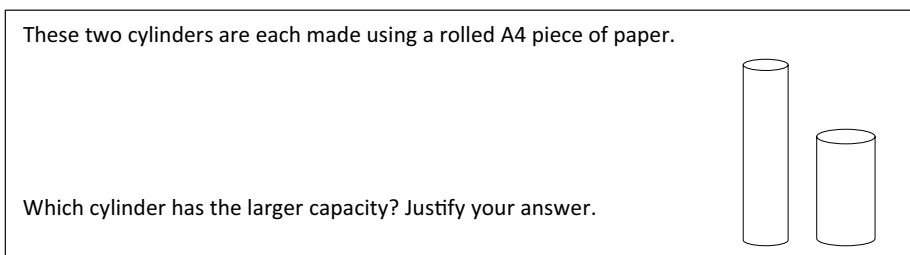


Fig. 3 Cylinder comparison task

and again. So I have a lot of squares, having the same capacity no matter how we roll them. Therefore, the capacity of the two cylinders is equal. (school student)

PST2 scored the response as ‘low’ for the criterion of ‘understanding the problem’, ‘low’ for correctness of final answer, and ‘medium’ for ‘creativity’. PST26, PST2’s feedback provider, wrote that the student’s solution was very creative and could serve as a basis for a further class inquiry. She also wrote that in contrast to PST2, she thought that the student *did* understand the problem, but used unjustified assumptions: that the fact ‘the sum of the areas of all rectangles equals the total area of the paper’ must also apply to the cylinders’ capacities. PST2 reported that initially she did not agree with PST26 but was convinced after their conversation that the solution was unique and that it was important to encourage students to try nonroutine solution paths. She therefore decided to score the student response as ‘high’ on creativity and to challenge him to further investigate his solution. When asked about her perceived learning gains, PST2 wrote:

The conversation with PST26 was a big thing for me. We did not agree on everything, but looking together at my assessment from different perspectives, and having to explain myself to her, and listening to her ideas, and thinking together [my role], I felt that we went deeper and deeper into the students’ thinking and the assessment they deserve [I believe]. (PST1)

PST2 reflects on how the process of collaborating with PST26 contributed to learning on their role as feedback providers, and at the same time, as an outcome of their collaboration, to their growth of knowledge on students’ mathematical thinking.

This example highlights the provocative role that peer assessment played in encouraging both PSTs in the pair to think more deeply about a rich task itself—the mathematics involved—and about knowing what to value in a student’s response, particularly when it does not match one’s expectations.

C3. Awareness of the importance of openness to students’ own ways of thinking Just over half of the PSTs expressed a perceived increase in their readiness to look at a student’s solution in a flexible manner as a salient learning outcome. Several PSTs reported that the dialogue with their feedback provider prompted them to learn to appreciate solutions that were different from the ones they had preconceived. This is suggestive of a shift in the PSTs’ assessment literacy, perhaps in terms of their overall disposition—*knowing* that students can show unexpected ways of thinking, *feeling* appreciative, and *believing* that such thinking is valuable. For example, PST6 wrote:

In the case of the heads and legs problem [of finding the numbers of chickens and dogs when given the total numbers of legs (78) and heads (25)], I was kind of expecting an algebraic solution. That was the solution I had in mind. I could not appreciate other solutions, like Amir’s, for example [see Fig. 4]. I measured his solution against my expectations and decided that his solution was not sophisticated [I believe]. The conversation with PST22, my colleague, opened my mind [I know] to see the beauty in Amir’s thinking. I think that I have learnt an important lesson: to read students’ answers with an open mind [I know] and not to be fixed firmly on my own approaches [I believe]. (PST6)

In her response, PST6 talks about the way that her collaboration with PST22 shaped her disposition towards assessing students’ thinking while valuing their unique contributions. Developing such an awareness that as an assessor, a teacher tends to bring preconceptions about appropriate responses to a task, can help teachers approach assessment

Fig. 4 Amir's solution to the heads and legs task



with more openness to students' diversity and mathematical ideas (Danielson and Marquez 2016). It is encouraging that PSTs were found to attend to this already at an early stage in their professional learning.

C4. Increased awareness of needing to align assessment goals, task, and rubric Just over half of the PSTs reflected on how the peer-assessment cycle helped them attend to the importance of having a match among assessment goals, the chosen mathematical task, and the rubric/criteria. This idea is related to the first theme of designing rubrics for rich tasks (C1), but in contrast to C1, it was not expressed by the PSTs as restricted to a specific assessment criterion as such, but rather as a meta-level principle to remember when developing an assessment. For example, during the whole-class discussion, PST15 (the student–teacher who had chosen the task on substitution of positive/negative numbers into an expression) reflected:

I realize now how much work is involved in planning an assessment activity [I know]. In choosing the right task or designing it so as to meet your goals and defining the criteria you use for assessment, they all have to support and mirror each other [my role]. (PST15)

Similarly, PST27 referred to her coming to appreciate a crucial component of fulfilling her role as assessor:

Something that I am taking with me is to think deeply about the connections among the mathematical skills and understandings I aim to assess [my role], the mathematics task that should focus on these skills and understandings, and the assessment criteria I use [I know]. Otherwise the assessment does not make sense. One of the criteria I used in my assessment, for example, was 'justification', which for me is fundamental in mathematics [I believe]. However, the task that I used did not explicitly ask for justification, and moreover, I did not define what is the nature of justification that I am expecting for, or in other words, what would count as a qualitative justification (deductive? And what about numerical examples?). So in that case assessing quality of justification was meaningless [I know]. (PST27)

These examples highlight that peer feedback on their assessment practice using a specific mathematics task can prompt PSTs to notice more general assessment principles they can apply in their future role as assessors to other assessment tasks.

Affective aspects of the PSTs' perceived experience

After participation in the assessment cycle of activities, the PSTs each gave scores for their perceived levels of confidence (1–5) both in assessing school students' mathematics work and in providing feedback to peers on their rubrics and student assessments. Their scores are presented as a cross-tabulation in Table 1.

As shown in Table 1, on average, the PSTs' expressed level of confidence in their role as assessors of students' mathematics work was lower than their level of confidence in their role as feedback providers to their peers. This seems suggestive of these PSTs' perception of the need for their development of student mathematics assessment expertise, also highlighted in previous research findings (Graham 2005; Volante and Fazio 2007). In addition, this contrast suggests that the PSTs were comfortable with the peer-assessment processes that were used formatively (rather than summatively) in this study. For example, one of the PSTs, PST18, wrote:

I think that I gave comprehensive feedback [I am confident]. I carefully went over the task, the criteria, the students' work and the teacher's evaluation. I indicated what I agreed with, what I disagreed with, and why [my role]. I enjoyed it [I feel] because I felt that I was doing a good job [I am confident] ... I think it was easier for me to respond to my colleague's assessment of student work than to do my own assessment of student work [I am/am not confident] ... Maybe because with being a "reader" I have to relate to something that already exists, which makes it easy for me [my role]. (PST18)

PST18's response reflects high confidence in her capability of providing feedback for her colleague's assessment ('I gave comprehensive feedback', 'I was doing a good job'). There is also evidence of her positive feelings in providing feedback ('I enjoyed it'). The high confidence she felt as a feedback provider is contrasted by her with less confidence in her capability to assess her own students' work ('it was easier for me to respond to my colleague's assessment of student work than to do my own assessment of student work'). She explains this discrepancy by referring to the nature of the role of a feedback provider which requires relating to 'something that already exists': a student assessment already made by a colleague.

Five themes emerged from analysis of the PSTs' reflections about their experiences with assessment practices in terms of affective dimensions and were evident in the group interview responses; they also seemed to be agreed on by the group as a whole. Two of them were related to self-perceived areas of confidence (A1 and A2; A stands for Affective), three themes were related to self-perceived areas of less confidence (A3, A4, and A5), and

Table 1 A cross-tabulation of PSTs' (by #) perceived levels of confidence (1–5; 5 being the highest) providing feedback to peers on their assessment (F) and assessing secondary students' mathematics work (A)

	A1	A2	A3	A4	A5
F1					
F2					
F3		#6	#8, #9		
F4		#10, #12, #20, #25	#1, #5, #13, #26	#3, #11, #15	
F5		#16, #21, #23, #24	#2, #7, #18, #27	#4, #14, #19	#17, #22

one theme was suggestive of a shift in beliefs about students' solutions to FA tasks and about the teacher's role (A6).

A1. Confidence in designing mathematical task and assessment criteria Nearly half of the PSTs expressed satisfaction from their abilities to design (or adapt) a suitable mathematics task and assessment criteria for it. For example, PST11 wrote:

I think that I was successful in building the assessment tools [I am confident]. I set myself goals for the assessment task, I chose a problem and then generated assessment criteria that fit the goals [my role]. For me the important assessment aspects were the process of solving the task, that is, the model required by the task and the strategies used, and students explaining their thinking [I believe]. Once I had set these goals, I knew what I was looking for [I know]. (PST11)

PST11 reflects on her sense of self-efficacy as an assessor. Her response highlights what she believes is associated with assessing students' work (setting goals for the assessment task, choosing a problem, and generating assessment criteria that fit the goals). Her confidence seems to be related to her systematic moving through the assessment process, which perhaps makes her feel competent in her practice, with specific attention to deciding on appropriate assessment criteria (the model, the strategies used, communication) that facilitates her evaluation of a student's work.

This suggests that several PSTs in this cohort did not find the challenge of creating rubrics for rich assessment tasks to be out of reach of their capability at this stage in their course.

A2. Confidence in attending to the details in students' responses Just over a third of the PSTs expressed satisfaction from their abilities to analyse student task responses deeply and pay attention to the details in the solutions to the problem. For example, PST13 wrote:

I was trying to look for possible sources of student errors and focused on reading the solutions in depth and attending to minor details [my role]. That is not to say that it was always easy, but I felt that I was able to do a good job [I am confident]. (PST13)

PST13 reports on a sense of self-efficacy associated with her assessment skills in attending to students' strategies and interpreting their understandings.

Several PSTs also mentioned that receiving feedback from their peers, which agreed with major aspects of their own student assessments, contributed to feeling positively about their own capability.

Three main areas of a lack of confidence were identified in the PSTs' reflections and emerged also in the group interview, when reflecting on their role as assessors of school students' mathematics task responses.

A3. Less confidence in choosing a task and designing appropriate assessment criteria Nearly half of the PSTs expressed confusion in choosing a suitable rich task and deciding on appropriate assessment criteria, in contrast to several other PSTs who had described these practices as an area of confidence. A key reason for these feelings seemed to be related to the PSTs' uncertainty in how to define a 'high-quality' solution for a chosen task. For example, PST20 wrote:

I found it difficult to define my expectations for a good solution to the problem [finding the perimeter of a chain of 10 hexagons represented in a geometric growing pattern] [I am not confident]. Should I accept any kind of strategy they use, like counting? Is an algebraic solution better? Is a recursive rule worse than an explicit rule? [I do not know].

PST20's response expresses a confusion in determining what counts as a solution of a good quality. His uncertainty seems to relate to a shortage of pedagogical content knowledge, that a symbolic generalization is regarded as the highest level of algebraic thinking and recursive approach to generalization is considered poorer than an explicit one (e.g. Radford 2006).

PST20 continued, raising an issue that was repeatedly raised by the participants—how to determine criteria for assessing students' creativity:

I also wondered about creativity. I wanted to creativity to be valued [I believe], but what counts as a creative solution to this task? [I am not confident] (PST20)

Another recurrent reason for feeling confusion in creating a rubric seemed related to PSTs' unfamiliarity with the range of students' possible solutions, misconceptions, or issues, thus making it hard for them to decide beforehand on suitable assessment criteria to assess diverse responses. For example, PST21 wrote:

For me a main difficulty was to anticipate students' answers and difficulties [I do not know]. I chose a rich task and wanted to define the assessment criteria that looked appropriate for the task: building a model, inferring conclusions, communication, and creativity [my role]. I tried to imagine what possible solutions students would give and what kinds of mistakes they might make so as to decide whether my criteria were appropriate [my role]. But it was difficult to think about solutions that are different from my own one [I do not know]. Only after I had received solutions from a group of students was I able to critically evaluate my criteria and revise them [I know]. Maybe when becoming a 'real' teacher, things will get better? [I believe] (PST21)

A4. Less confidence in grappling with the subjective nature of certain assessment criteria Two-thirds of the PSTs expressed confusion, and even anxiety, related to evaluating the quality of students' solutions, and particularly the dimensions of *creativity* and *communication*, although occasionally with respect to other criteria as well. Several PSTs expressed their belief that developing creativity is an important part of mathematics learning and that they see creativity as valuable to assess. In practice, however, when evaluating the extent of creativity of a certain solution, it was challenging. They described a feeling of hesitation about their 'right' to judge creativity as it seemed to be subjective. For example, PST6 wrote:

I found it very difficult to assess a solution for its creativity [I am not confident]. For example, I thought that Amir's numeric solution (see Fig. 4) for the heads and legs problem was not creative [I believe] as it did not show a unique or exceptional solution method. However, PST22 (my peer) said that precisely because it is not the usual algebraic solution, it is unique and creative. So who is right? [I do not know] (PST6)

Other PSTs also reflected on their confusion when trying to assess *communication*. A common concern was whether or not a teacher should expect a descriptive explanation for a solution that 'speaks for itself'. For example, PST3 wrote:

When I tried to assess the students' responses for their quality of communication, I felt confused [I feel]: There were cases in which I could not understand a student's intentions because the solution lacked an explanation. In such cases explanations are required in order to get a high mark in communication [I know]. But there were other cases in which the solution 'spoke for itself' and I did not

need further explanation in order to understand the underlying thinking in it. So should I deduct points for not including an explanation, or not? [I am not confident] (PST3)

She then gave an example:

For example, a student drew three parabolas as an answer to the task [see Fig. 5] for the actual task PST3 used in her assessment and the student's solution]. I like the solution, and I think it reflects understanding [I believe], but I still wonder, should I require a descriptive explanation from him as well? Can I say that the solution is perfect in terms of communication [I do not know]?

Similarly, PST's reflection, already presented in the data analysis section, expresses her anxiety related to evaluating the quality of students' solutions:

I have learnt to select good rich tasks and to build appropriate criteria for assessment [I know]. However, something bothers me [I feel]. When sharing my assessment with my colleague, I found it very confusing [I feel] that one of us evaluated a solution as excellent, whereas the other assessed it as a medium level of quality. It made me feel uncomfortable and nervous [I feel]. What will I do when I become a real teacher? How will I know that I have interpreted the student's response correctly? [I feel]. (PST8)

A5. Less confidence in interpreting students' thinking Half of the PSTs expressed confusion for struggling in interpreting and evaluating students' ways of thinking based on their written solutions. This was difficult, they claimed, especially because the solutions were for a rich task, and students did not always offer clear explanations of their solution process. For example, PST26 wrote:

What was difficult for me sometimes was being sure about my understanding of a student's way of thinking based on reading his answer [I am not confident]. The task is open ended, and the solution is not obvious in its presentation or its reading, and this makes it hard to assess. Indeed, there was a case where my peer and I interpreted a student's solution differently. The student had presented her solution and had obtained the wrong answer. The explanation was very vague. I thought she had made a calculation error, but my colleague thought that the student had totally misunderstood the task questions. (PST26)

$f(x)$ is a quadratic function. Given that $f(1) = 10$ and $f(6) = 6$, how many solutions does the equation $f(x) = 0$ have? Explain in as much detail as you can in your response.

The student's solution:

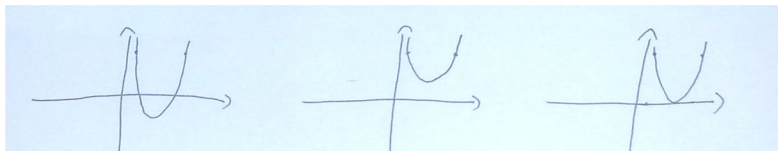


Fig. 5 Task and a student's solution, as given by PST11

This difficulty might be more related to school students' lack of experience in presenting written solutions to rich tasks than to the PSTs' assessment struggles and is likely assuaged by students learning to attend to their communication in rich task responses.

A6. Development of humility in responding to students' work Nearly half of the PSTs reported on developing a sense of humility when assessing students' knowledge and understandings: an awareness that teachers need to keep in mind the limitations of their own interpretations of students' solutions. At different times, it would be wise for teachers to test their personal interpretations with their teacher colleagues or further with the students themselves. The PSTs described two main reasons leading them to consider humility as an essential value within their assessment practice. One reason was related to the multiplicity of voices they encountered during the peer-assessment process. For example, Kathrin wrote:

I learnt that different teachers may comprehend students' solutions differently [I know]. This means that my own view about a student's level of understanding is only one view among others. I think it is good for us to remember that our thinking is not the only 'truth' and that it is only one possible interpretation of the student's knowledge [I believe]. Therefore, in order to provide the student with a fair score and appropriate comments, sometimes it is better not to rely solely on my view as is, but to ask the student for further explanations about his answer [I believe]. In this way my own understanding of his knowledge will become closer to 'the truth' [my role]. (PST7)

Likewise, PST6 (referring again to her initial expectation of an algebraic model being the only appropriate response) wrote:

It was an important experience for me to become aware of the subjective nature of an assessment, which I guess is formed and restricted by our experience and expectations of students [I believe]. So it is important for me and all other teachers to be aware of the possibility that our opinion is not necessarily complete [I believe]. (PST6)

Another aspect related to the theme on developing humility was the difficulty the PSTs sometimes encountered in judging student understanding based on a student's written solution. Occasionally, students' written answers were not well explained or clear; thus, their thinking may not be obvious. For example, PST9 wrote:

We do not always know for sure what the student was thinking [I am not confident]. Maybe he thought correctly but did not express himself well. This is why we should decrease our degree of confidence while assessing so as to be fair with the students [I believe]. (PST9)

These examples illustrate the PSTs' attitudinal shift towards valuing humility, fairness, honesty, and responsibility when assessing—to keep an assessment task in proportion by acknowledging that a student's task response is evidence of, but does not fully capture, that student's knowledge and understandings. It seems that encountering differing views about student task responses from peers, and having the opportunity to discuss them, highlighted to the PSTs the value of being more nuanced and open in their approach to making judgements about students' thinking.

Discussion and conclusion

This study sought insights into the perceived experiences of secondary mathematics prospective teachers in learning through authentic approximations of formative assessment practice (Grossman et al. 2009). In a peer-assessment cycle, they selected rich mathematical tasks, designed rubrics with assessment criteria, marked (real) school students' task responses with their rubrics, and both gave and received formative peer feedback. The study focused on the PSTs' perceptions of their experience of formative peer-assessment processes for learning to assess school students, in terms of cognitive and affective dimensions of their conceptions of assessment. In this study, the PSTs described learning outcomes salient to them and four cognitive-related themes emerged from the analysis, suggestive of increasing their appreciation of student diversity and of increasing their awareness of the importance of careful assessment task design. It seemed that the PSTs linked these outcomes to their discussion and negotiation with peers about the rubrics and assessments, rather than simply being about having further experience with assessment practices per se. One theme appeared to be related more to global/meta-level learning about assessment practice and was referred to by half of the PSTs. It was that giving and receiving peer feedback prompted them to notice misalignment among assessment goals, the chosen mathematical task, and the rubric/criteria—which is considered critical in effective formative assessment (Swan and Burkhardt 2012). Through discussion, they attended more reflexively to their practice in designing and assessing tasks. The literature highlights the importance of consultation during the assessment task development process (Quinlan 2006; Stevens and Levi 2005). This study's findings highlight that PSTs experiencing this consultation for themselves through formative peer assessment may become more attuned to this important principle.

The study found that after their experiences, the PSTs on average expressed a medium level of confidence for assessing students' rich task responses. Analysis of their reflections revealed that the cohort was evenly divided between those who expressed confidence in their assessment practices and those who described a lack of confidence with those same practices: the construction of a rubric for a selected rich task and in its subsequent use to assess students' solutions. For several PSTs, who had initially felt confident about their assessment work, the sense of complexity associated with assessing students' thinking only surfaced later when receiving feedback on their assessment from their peers. They described finding out from a peer that the criteria they constructed—which initially appeared to them to work well when using them to assess students' work—were incomplete or inappropriate and needed revision. This suggests that the feedback from a peer on their assessment efforts helped them attend to problematic aspects or 'blind spots' of their knowledge. This resonates with similar findings about practising teachers, whose collaborative marking of rich tasks highlighted issues such as ambiguous wording in the rubric or differing interpretations of criteria (Danielson and Marquez 2016).

In terms of choosing a rich task and designing assessment criteria for it, the analysis revealed two main sources of several PSTs' expressed lack of confidence. One was related to their sense of uncertainty in defining expectations as to what counts as a high-quality solution for a particular task—what ought to be valued by the teacher. The second was related to their lack of familiarity with the range of students' possible solutions, misconceptions or issues, thus making it hard for PSTs to decide beforehand on suitable assessment criteria to cover diverse responses. Wilkie (2019) found that secondary mathematics teachers described rich (pattern generalization) tasks as 'valuable' but 'complex' to assess

since the process was not like usual tests involving ‘tick, cross, 1 mark, 2 marks’ (p. 20). As for using their constructed rubric to analyse actual student solutions, again two main sources of PSTs’ difficulties were described by them. One was related to their feeling of ambiguity in trying to judge the quality of a rich task solution, particularly in terms of its *creativity* (‘too subjective’) or *communication* (‘should I require a descriptive explanation?’) (Evans and Swan 2014; Schwartz et al. 1995). The second was related to struggling to understanding a student’s intentions and thinking, which could be ‘hidden’ in a vague solution. According to the PSTs’ reports, dialogue with their peer feedback providers enabled them to develop more flexibility in interpreting students’ solutions and appreciating solutions different from the ones they originally had in mind.

Such challenges in assessing rich tasks may be exacerbated due to PSTs being less experienced in assessment practice than practising teachers and less knowledgeable about students and curriculum in ways that would support their use of formative assessment (Heritage 2007). Nevertheless, as highlighted by the current study’s participants, their experience of the peer-assessment cycle—wearing the ‘hats’ of assessors of students’ mathematics for themselves and also feedback providers on peers’ assessment practice—enabled them to some extent to notice and grapple with these challenges in a reflexive way. More particularly, the PSTs reported that receiving peer feedback on their criteria and judgements, and negotiating verbally during the peer conference, encouraged them to think more deeply about and reconsider their assessment practice through hearing differing perspectives. Comments and suggestions for removing, adding, or changing the weighting of a certain criterion offered by their peers appeared to help them notice gaps in their assessment and distinctions between levels of solution quality.

The literature on undergraduate peer assessment experiences have highlighted that peers can be reluctant both to critique the work of others (Hanrahan and Isaacs 2001) and to receive such critiques from one another (Smith et al. 2002). Surprisingly, this study did not find that the PSTs expressed any negative sentiments about this in their reflections. They did refer to disagreements but couched them in positive terms, as benefiting their learning. Perhaps the wording of the reflection prompts encouraged a focus on learning outcomes. We think it is probable that in using peer assessment for learning, rather than for contributing to a summative grade, the potential for negative affect or power issues (van Gennip et al. 2010; Topping 2010) was alleviated. Huberman (1995) in discussing practising teachers’ interaction emphasized that even though some level of mutual affective comfort between teachers is necessary, it is likely to be successful in encouraging reflection on improving practice, if each teacher is committed to the work in which they are engaged. They do not need to agree, but need to be open to another’s perspective so to try to understand it as much as possible.

On average, the PSTs reported a higher level of confidence when providing feedback on their peers’ assessment of students’ work than when assessing school student work themselves. They indicated that they felt capable of evaluating critically and providing effective comments to improve a peer’s assessment work. Studies have shown that assessing one’s own work is often more difficult than judging the work of others because the assessor is ‘too close’ to the work, lacking a distanced, objective perspective (Black et al. 2003; Dunlosky and Lipko 2007). Peer assessment appears to allow students to step back in a similar manner and allow ‘working through’ to be replaced by ‘working on’ (van Gennip et al. 2010). In addition, peer feedback can be beneficial to learning due to the absence of a clear ‘knowledge authority’ and the uncertainty induced by a peer’s relative status that may encourage students to search for confirmation, leading to further thinking and discussion (Yang et al. 2006).

Nearly half of the PSTs reported on developing a sense of humility when making judgments about students' knowledge and understandings, to not presume that there is nothing further to learn about a student's thinking. According to the PSTs' reflections, they developed more awareness about their current limited understanding of students' solutions and that students' thinking may be more sophisticated than they first thought. They expressed their perceived value of teachers being open to soliciting the feedback of colleagues and students themselves when assessing these types of tasks and being ready to inquire further into students' thinking when unclear in a written task response. Large-scale longitudinal research found that teachers' views of student capabilities shaped the extent to which they elicited and built on a wide range of student thinking (Cobb et al. 2018). Teachers holding deficit views of student thinking (in attributing difficulties to inherent student characteristics) were found to be less likely to adopt reform teaching practices, such as rigorous problem solving. A teacher's humble attitude—in recognizing one's own biases, limitations, and lack of knowledge—is related to teaching practices that encourage student participation (Sezgin and Erdoğan 2018). Such knowledge, beliefs, and expectations, which vary with an individual teacher's personal, social, and cultural history, influence a teacher's assessment of students' work (Morgan and Watson 2002). The multiplicity of voices the PSTs encountered in this study during the peer-assessment process seemed to lead them to consider humility as an essential stance in assessment practice. Future research on PSTs' views, in the context of assessment, would be worthwhile, but this study's findings highlight the potential benefit of peer feedback for disrupting or at least challenging existing preconceptions or deficit attitudes.

Exploration of the perspectives of this cohort of PSTs, even though for a short duration with one peer-assessment cycle, enabled us to consider some of the likely advantages and challenges associated with applying peer assessment as a learning tool in teacher preparation courses. The emergent themes will support more extensive research on strategies to prepare secondary mathematics prospective teachers effectively for utilising formative assessment principles and practices. This study did not investigate PSTs' subsequent implementation of formative assessment in the classroom practice. Yet an initial step to encouraging teacher enactment is through engaging teachers in situations that prompt them to re-examine events and come to view them differently (Kennedy 2016). In the case of this study, the peer-assessment cycle seemed to support PSTs in seeing formative assessment practices and their role, knowledge, and values in different and more complex ways. Further research is needed to explore the ways in which participation in such professional learning may be realized in classrooms. Additionally, this study, as far as we know, is an initiative exploring the potential of peer-assessment cycles with PSTs for learning Pedagogical Content Knowledge (Shulman 1986). The findings of the study, in terms of the PSTs' perceived learning gains, provide some evidence as to the likely success of such a strategy. Other researchers may adapt the peer-assessment cycle used in this study for enhancing teacher learning of practices other than formative assessment.

In drawing on the TALip model (Xu and Brown 2016) for researching pre-service teacher assessment literacy development, we found that cognitive and affective aspects appear intertwined. For example, surprise about a mismatch between a prior (poor) opinion of a student solution and a peer's (positive) interpretation of it seemed to prompt awareness of a lack of knowledge about mathematical strategies, and an attitudinal shift towards humility. We grappled with differing ideas in the literature about defining the cognitive and affective dimensions, particularly when analysing the PSTs' expressions of belief. These differences contribute to an ongoing conversation among researchers who study teacher learning (Philipp 2007). Yet we found explicit attention to both dimensions can give a rich

picture of teacher assessment literacy development and resonated with the participants themselves in sharing their reflections. In this article, we did not investigate these relationships in a systematic way but they deserve further research.

Appendix 1: Illustrative rubrics used/adapted by the PSTs

Illustration #1 (adapted from Schwartz et al. 1995)

Assessment criteria	High	Medium	Low	Comments
Modelling: How well does the student build the model for solving the mathematical problem?				
Manipulating: How well does the student manipulate the mathematical formalism in which the problem is expressed?				
Inferring: How well does the student apply the results of his or her manipulation of the formalism to the problem situation that spawned the problem?				
Communicating: How well do students communicate to others what they have done in formulating the problem, manipulating the formalism, and drawing conclusions about the implications of their results?				
Creativity: How much original is the student's solution?				

Illustration #2 (another version from Schwartz et al. 1995)

Assessment criteria	High	Medium	Low	Comments
Understanding the task (e.g. the situation, instructions, data, restrictions)				
Mathematical thinking (e.g. generalizations, justifications)				
Manipulations				
Communication (e.g. explanations, presentation of the solution, mathematical language)				
Creativity (originality and multiple solutions)				

Illustration #3 (built on Swan and Burkhardt 2012)

Assessment criteria	High	Medium	Low	Comments
Representing (selecting information, methods and tools)				
Analysing (e.g. making connections, making calculations, making conjectures and generalizations)				
Interpreting and evaluating (forming conclusions and arguments, considering appropriateness and accuracy, relating back to the original situation)				
Communicating and reflecting (e.g. communicating and discussing findings effectively, considering alternative solutions)				

Appendix 2: Feedback provision by writing an analysis for a peer

Imagine that you are sitting in a team meeting of the mathematics teachers at your school. One of the teachers shares with you the rubric he/she has built to assess student solutions for a certain task. He/she asks you for your opinion on the rubric he/she developed and on the assessment he/she gave with the help of the rubric to the students' solutions.

1. Read the task, the rubric, the student solutions, and your peer's student assessments in the report (see the peer's name below the work to which you are to respond).
2. Give feedback to your peer:
 - a. What is good and what would you improve on the rubric (in accordance with the task chosen)?
 - b. What do you agree with and what aspects would you change in their assessment of student solutions with the rubric? Justify your comments so that your peer can understand them.

In continuation of the 'conversation' with your colleague at the teachers' meeting at school, imagine your peer asked for your advice as to what response you would give to each of their students in order to advance his/her understanding and learning.

3. Offer your peer responses to each student according to his/her task solution. Explain and justify your suggestions.

Appendix 3: Coding scheme with theme frequencies

Coding scheme		Num. PSTs (<i>n</i> = 27)
<i>Cognitive aspects</i>		
C1	Increased knowledge about designing rubrics for rich tasks	20
C2	Knowledge of how to distinguish between levels of quality in responses	18
C3	Awareness of the importance of openness to students' own ways of thinking	16
C4	Increased awareness of needing to align assessment goals, task, and rubric	13
<i>Affective aspects</i>		
A1	Confidence in designing mathematical task and assessment criteria	12
A2	Confidence in attending to the details in students' responses	10
A3	Less confidence in choosing a task and designing appropriate assessment criteria	13
A4	Less confidence in grappling with the subjective nature of certain assessment criteria	18
A5	Less confidence in interpreting students' thinking	14
A6	Development of humility in responding to students' work	13

References

- Andrade, H. L., & Cizek, G. J. (2010). *Handbook of formative assessment*. New York, NY: Routledge.
- Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, 60(5), 497–511.
- Black, P. (2015). Formative assessment—An optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, 22(1), 161–177.
- Black, P., Harrison, C., Lee, C., Marshal, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The Sage handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks, CA: Sage.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1–14.
- Chick, H., & Beswick, K. (2017). Teaching teachers to teach Boris: A framework for mathematics teacher educator pedagogical content knowledge. *Journal of Mathematics Teacher Education*, 21, 475–499.
- Cobb, P., Jackson, K., Henrick, E., Smith, T. M., & MIST team. (2018). *Systems for instructional improvement: Creating coherence from the classroom to the district office*. Cambridge, MA: Harvard Education Press.
- Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., et al. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education*, 22(1), 3–29.
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Crossman, J. (2007). The role of relationships and emotions in student perceptions of learning and assessment. *Higher Education Research and Development*, 26(3), 313–327.
- Danielson, C., & Marquez, E. (2016). *Performance tasks and rubrics for high school mathematics: Meeting rigorous standards and assessments* (2nd ed.). New York, NY: Routledge.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419–438.
- Dunlosky, J., & Lipko, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, 16, 228–232.
- Evans, S., & Ayalon, M. (2016). Can designed student responses support teachers to interact with students in a productive way? *Educational Designer*, 3(9), 12.
- Evans, S., & Swan, M. (2014). Developing students' strategies for problem solving. *Educational Design*, 2(7). Retrieved from <http://www.educationdesigner.org/ed/volume2/issue7/article25/>. Accessed 25 May 2017.
- Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, 27(4), 403–434.
- Ghousseini, H., & Herbst, P. (2016). Pedagogies of practice and opportunities to learn about classroom mathematics discussions. *Journal of Mathematics Teacher Education*, 19, 79–103.
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21, 607–621.
- Green, T. F. (1971). *The activities of teaching*. New York: McGraw-Hill.
- Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice*, 15(2), 273–289.
- Hanrahan, S., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research & Development*, 20(1), 53–70.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140–145.
- Huberman, M. (1995). Networks that alter teaching: Conceptualizations, exchanges and experiments. *Teachers and Teaching*, 1(2), 193–211.
- Kennedy, M. (2016). Parsing the practice of teaching. *Journal of Teacher Education*, 67(1), 6–17.
- Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, 20(4), 344–348.

- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442–467.
- MacLellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education*, 20, 523–535.
- McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575–596). Charlotte, NC: National Council of Teachers of Mathematics, Information Age Publishing.
- Merseth, K. K. (1991). *The case for cases in teacher education*. Washington, DC: American Association for Higher Education, American Association of Colleges of Teacher Education.
- Morgan, C., & Watson, A. (2002). The interpretative nature of teacher's assessment of students' mathematics: Issues for equity. *Journal for Research in Mathematics Education*, 33(2), 78–111.
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- Patton, C. (2012). 'Some kind of weird, evil experiment': Student perceptions of peer assessment. *Assessment & Evaluation in Higher Education*, 37(6), 719–731.
- Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 257–315). Charlotte, NC: National Council of Teachers of Mathematics, Information Age Publishing.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11.
- Quinlan, A. (2006). *Assessment made easy: Scoring rubrics for teachers from K–college*. Lanham, MD: Rowman and Littlefield Education.
- Radford, L. (2006). Algebraic thinking and the generalization of patterns: A semiotic perspective. In *Proceedings of the 28th conference of the international group for the psychology of mathematics education*, Chapter. Mérida: Universidad Pedagógica Nacional (Vol. 1, pp. 2–21).
- Reinholz, D. (2016). The assessment cycle: A model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 41(2), 301–315.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77–84.
- Schafer, W. D., Swanson, G., Bené, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education*, 14(2), 151–170.
- Schwartz, J., Kenney, J., Kelly, K., Sienkiewicz, T., Sivan, Y., Steinbok, V., & Yerushalmy, M. (1995). *Assessing mathematical understanding and skills effectively*. President and Fellows of Harvard College.
- Sezgin, F., & Erdoğan, O. (2018). Humility and forgiveness as predictors of teacher self-efficiency. *Educational Research and Reviews*, 13(4), 120–128.
- Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, 63, 66–70.
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Martens, R. (2004). Training teachers in peer-assessment skills: Effects on performance and perceptions. *Innovations in Education and Training International*, 41, 59–78.
- Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: A case for student and staff development. *Innovations in Education and Teaching International*, 39, 71–81.
- Stevens, D. D., & Levi, A. J. (2005). Leveling the field: Using rubrics to achieve greater equity in teaching and assessment. *Essays on teaching excellence, Professional and organizational development Network in Higher Education*, 17(1), 1–6.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534–539.
- Swan, M., & Burkhardt, H. (2012). A designer speaks: Designing assessment of performance in mathematics. *Educational Designer: Journal of the International Society for Design and Development in Education*, 2(5), 1–41.
- Taras, M. (2007). Assessment for learning: Understanding theory to improve practice. *Journal of Further and Higher Education*, 31(4), 363–371.
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249–276.
- 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, NY: Routledge.
- Topping, K. J., & Ehly, S. (Eds.). (1998). *Peer-assisted learning*. Mahwah, NJ: Erlbaum.

- van Gennip, N., Segers, M. S. R., & Tilema, H. H. (2010). Peer assessment as a collaborative learning trajectory: The role of interpersonal variables and conceptions. *Learning and Instruction, 20*, 20–290.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education, 30*(3), 749–770.
- Walkoe, J. (2015). Exploring teacher noticing of student algebraic thinking in a video club. *Journal of Mathematics Teacher Education, 18*(6), 523–550.
- Wilkie, K. J. (2019). The challenge of changing teaching: Investigating the interplay of external and internal influences during professional learning with secondary mathematics teachers. *Journal of Mathematics Teacher Education, 22*(1), 95–124.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy and Practice, 11*(1), 49–65.
- Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Lawrence Erlbaum Associates.
- Xu, Y., & Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149–162.
- Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*, 179–200.
- Zevenbergen, R. (2001). Peer assessment of student constructed posters: Assessment alternatives in pre-service mathematics education. *Journal of Mathematics Teacher Education, 4*, 95–113.

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