FACULTY OF PSYCHOLOGY
AND EDUCATIONAL SCIENCES

The social nature of peer assessment in secondary and higher education

Examining students' perceptions on interpersonal processes and peer feedback quality in anonymous face-to-face settings using Mobile Response Technology

Tijs Rotsaert

Supervisor: Prof. Dr. Tammy Schellens

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Doctor of Educational Sciences

Academic year 2016–2017

GHENT
UNIVERSITY

VOORWOORD

Vier jaar geleden begon ik aan de uitdagende opdracht om, in een onderzoeksveld dat in volle groei en beweging is, een doctoraat over *peer assessment* te maken. Aan het einde van deze periode kan ik jullie met trots mijn proefschrift voorleggen. Het boek dat hier nu ligt was er echter niet geweest zonder de steun van heel wat mensen. Enkele van hen wil ik hier in het bijzonder bedanken.

Eerst en vooral wil ik mijn promotor, **prof. dr. Tammy Schellens** bedanken. Tammy, bedankt voor de kans om dit project in jouw onderzoeksgroep uit te voeren. Dankjewel voor de vrijheid die ik kreeg om diverse onderzoekspistes en -settings te ontdekken en voor de ruimte om met collega's in het onderzoeksgebied samen te werken. Je merkte al gauw dat de teamplayer in mij af en toe naar buiten moest kunnen komen. De ontspannen overlegmomenten waarin vele keuzes besproken werden, zorgden ervoor dat ik steeds gesteund en gesterkt een volgende fase kon aanvatten.

'Muchas gracias' to my research partner *in crime* and dear friend, **dr. Ernesto Panadero**. Thank you, Ernesto, for the exciting collaboration. Your guidance during and after my research stay in the beautiful city of Madrid was enlightening and extremely valuable. Our common interest in PA led to fruitful explorations and new insights.

Verder wens ik ook de **leden van de begeleidingscommissie** – prof. dr. Bram De Wever, prof. dr. David Boud, prof. dr. Jan-Willem Strijbos en prof. dr. Katrien Struyven – te bedanken voor hun aandacht, tijd en suggesties tijdens de verschillende bijeenkomsten.

Ook een 'dikke merci' aan alle **collega's aan de vakgroep Onderwijskunde** voor jullie collegialiteit, ondersteuning en vriendschap. Jullie enthousiasme en expertise zorgt voor een heel boeiende werkomgeving. Een speciale dankjewel aan mijn 'meter' Annelies en aan mijn mede-eilandbewoners Britt en Lore voor het delen van lief en leed. De plantjes in ons bureau zullen het niet kunnen navertellen, maar de sfeer zat altijd goed! Top, dat ik in 'eer en geweten' de vele mannelijke clichés mocht doorbreken. Dank ook aan Hanne en Michiel – jullie maakten de sprint naar de eindmeet een heel stuk aangenamer.

Ik wil ook **alle leerkrachten en studenten** bedanken die meewerkten aan de verschillende studies. Jullie deelname was van onschatbare waarde! Dank aan Rik & Dirk voor de kans om het vak *Brede School* te mogen verrijken met nieuwe assessmentbenaderingen. Jullie visie en realiteitszin zorgden voor een vlotte vertaling naar de praktijk. Dank ook aan Dagmar voor het nauwkeurige codeerwerk.

Dankjewel aan de **leden van de SIG1-community en het VPO-bestuur**. Het was erg fijn om, naast het onderzoek, de analyses en het schrijfwerk, ook conferenties, workshops, een summer school en vooral leuke feestjes te kunnen organiseren. Ernesto en Jan-Willem, we maakten het ons niet gemakkelijk door de SIG1-conferentie in Munchen te organiseren vanuit Groningen, Madrid en Gent, maar geef toe: het resultaat mocht er zijn!

Merci aan de **vrienden van *de bende van de bosklapper* en de vele andere copains & copines** voor de oprechte steun en betrokkenheid, maar ook voor de relativering, de Cuba Libres, de heerlijke ontspannende avonden en weekends! Jullie weten dat jullie me heel nauw aan het hart liggen en ik kijk ernaar uit om tijd in te halen met jullie! Be prepared.

Tot slot nog een bijzondere dank voor mijn **ouders, oma en familie**. Lieve mama & papa, dat ik op dit punt ben geraakt, heb ik voor een groot deel aan jullie te danken. Bedankt voor de warme thuis, jullie onvoorwaardelijke steun en inspirerende kijk op het leven. Oma, dankjewel om mee te volgen vanop de eerste rij. Heerlijk om de laatste maanden in het weekend van jou de vraag te krijgen of ik nu al iets gehoord had van 'de reviewers'… Sanne en alle andere familieleden, ook aan jullie een dankjewel voor de bemoedigende 'gij kunt dats'.

Lieve **Linde**, de voorbije vier jaar waren we niet alleen verbonden in de liefde, maar ook in onze job. We zijn een onverwoestbare tandem. Onze nieuwe 'home office' in Ledeberg werd de voorbije maanden een inspirerende ruimte voor mezelf – en ik beloof je plechtig heel binnenkort mijn eerste Claus te lezen. Dankjewel voor alle begrip en steun maar vooral voor jouw onvoorwaardelijke liefde. So, what's next? ☺

Ledeberg, maart 2017

Tijs Rotsaert

TABLE OF CONTENTS

1 General introduction

Chapter 1
General introduction

"True intuitive expertise is learned from prolonged experience
with good feedback."

-Daniel Kahneman-

Abstract

This chapter provides a general overview of the research theme and the different studies that are presented in chapters 2 to 6 and discussed in chapter 7. After a general introduction, peer assessment (PA) is positioned as a potentially formative and sustainable assessment method. Subsequently, a conceptual framework is explained that explores three interconnected dimensions that are central to the enhancement of valuable peer feedback dialogue practices. These dimensions are: (a), the *social and interpersonal nature of PA*; (b), the *quality of peer feedback (PF)*; and (c), the *organization and management of PA practices*. The framework is used to identify possible barriers to and challenges for both research and practice within these three dimensions. The introduction concludes with the research objectives of this dissertation, an overview of the methodological approaches applied in each study, and an outline of the dissertation structure.

Introduction

Current approaches to assessment have moved away from end-of-course (i.e., summative) testing to in-course (i.e., formative), learning-oriented interactions between learners and instructors (Black & Wiliam, 1998; Havnes, Smith, Dysthe & Ludvigsen, 2012). This is conceptualized by Birenbaum (2003) as a shift from a testing culture to an assessment culture. The driving of the learning process by assessment appears to be most beneficial when we make students less dependent of the assessment process itself, and instead make them more responsible for their own learning. Active involvement in those processes is necessary, making assessment an integral aspect of learning (Havnes et al., 2012; Hunt, Hughes & Rowe, 2002). This shift blurs the

distinction between instruction and assessment (Cowie, Moreland & Otrel-Cass, 2013) and has become known as the assessment for learning (AfL) position (Black & Wiliam, 1998). Involving students in assessment processes contributes to their empowerment, and should provide them with necessary assessment skills for their professional development and lifelong learning (Planas-Lladó et al., 2014). It is up to educational research to explore the effectiveness of concrete interventions that may leverage the assessment literacy of students (Carless, 2013; Havnes et al., 2012).

Developments in the field predominantly stress the importance of formative approaches, where the information that has been gathered through the assessment task provides feedback to the students that is intended to steer their learning process, and can also be used by teachers to adapt the learning environment to better meet learners' needs (Wiliam, 2011). A key element within formative assessment processes is the provision of feedback to close the gap between current and desired performance. Feedback is described by Winne and Butler (1994) as "information with which a learner can confirm, add to, and overwrite, tune, or restructure information in memory, whether that information is domain knowledge, meta-cognitive knowledge, beliefs about self and tasks, or cognitive tactics and strategies" (p. 5740). In other words, feedback elicits elaboration and organizational cognitive processes, while restructuring one's schemata after incorporating feedback (Butler & Winne, 1995; Derry, 1996). In order to be effective, feedback should tackle three questions, each of which is formative and self-regulatory in nature (Clark, 2012): (a), feed-up: Where are we going? This concerns itself with the sharing of learning objectives; (b), feedback: How are we doing? A question which monitors and assesses learning progression, either for a specific task or more generally; and (c), feed-forward: Where to next? This question relates to the next steps required for improvement in a specific task/project, or more generally across time (Hattie & Timperley, 2007).

Parallel to and building upon the theory of formative assessment, a number of authors (e.g., Fastré, van der Klink, Sluijsmans & van Merriënboer, 2013; Boud & Soler, 2015) have taken up the idea of sustainable assessment. For these authors, formative assessment is promising in theory but on its own, is insufficient to help students learn to assess their own learning, and does not prepare them to act as lifelong learners beyond graduation and throughout their career. As such, the objective of sustainable assessment is entirely consonant with Bandura's (1997) perspective on the purpose of education: "to equip students with the self-regulatory capabilities that enable them to educate themselves" (p. 174). This approach takes a practice view that sees assessment as a key element of the process of developing and informing the learner's judgement for learning beyond the immediate task (Boud, 2007). The notion of sustainable assessment was developed to focus on the need for all assessment practices to equip learners for the challenges of learning and

practice they will face once their current episode of learning is complete. Boud suggests that "for students to become effective lifelong learners, they need also to be prepared to undertake assessment of the tasks they face throughout their lives" (2000, p. 152). Researchers in this area argue for a need to re-cast students as active agents in learning and consistent with this, to become active constructors of feedback information, generating it themselves and seeking it out from multiple sources, including the teacher. For students to learn from feedback, they must actively engage with the information; i.e., analyze the message, ask questions about it, discuss it with others, connect it with prior understanding, and use this to change future actions (Nicol, 2010). Within sustainable assessment practices, feedback becomes a process used by learners to facilitate their own learning (Boud & Soler, 2015). Building on these strands, feedback then involves a dialogic process, whereby learners make sense of information from various sources. Feedback dialogue takes a central place in this; "dialogue" here should not be taken literally as face-to-face, but rather all forms of interaction with different actors. Dialogue involves relationships in which participants think and reason together (Yang & Carless, 2013). As a consequence, teachers, policymakers and educational researchers are challenged to promote, design, and implement sustainable feedback practices with the following characteristics: a), involving students in dialogues about their learning; b), facilitating feedback processes through which students are stimulated to develop evaluative expertise; and c), assessment task design that facilitates engagement over time in which feedback from various sources is generated, processed, and used to enhance performance in multiple stages of assignments (Carless, 2015).

The theoretical shift described in the above paragraph has led to increased interest in new formats and tools of assessment (e.g., peer assessment, self-assessment, and computer-supported collaborative assessment). These AfL formats are becoming more widely adopted in classrooms, given that they are meant to support a better integration of teaching/instruction with assessment to monitor learning and provide feedback to modify learning, more than to establish accomplishments against predefined standards, as is typical in an "assessment of learning" approach (Birenbaum, 2003; Brown, 2004).

Peer assessment as a formative and sustainable assessment practice

Because of the active involvement of students within the assessment process, PA has been embraced as an innovative method of formative assessment and is often seen as a way to offer significant educational value for learning (e.g., Topping, 2010). PA is defined as any educational arrangement in which students judge their peers' performance by providing grades and/or offering written or oral feedback (Topping, 1998). Others would consider it to be a form of collaborative learning (Kollar & Fischer, 2010; Strijbos, Ochoa, Sluijsmans, Segers & Tillema,

2009). PA occurs more productively among individuals who have experienced a common learning context (Topping, 1998). In view of formative assessment, in the last decade, PA research has evolved – and is still evolving – towards practices that contain a component that includes giving and receiving PF, rather than only focusing on PA as a grading activity (e.g., Tseng & Tsai, 2007). In that sense, giving and receiving feedback from peers is seen as the most important learning element of PA (Liu & Carless, 2006). This view of PA is, therefore, strongly supported throughout this dissertation.

PA contains many benefits (Dochy, Segers & Buehl, 1999; Panadero & Brown, 2017; Topping, 2003) and can be conceived of as a learning tool due to the active involvement of the learner in the learning process and its provision of the learner with skills to assess criteria that define high-quality work (Orsmond, Merry & Callaghan, 2004; Topping, 1998). Peer assessment as an assessment tool is a strong vehicle of AfL because it actively involves students in evaluating their learning and allows them to participate in a collaborative appraisal using multiple perspectives, incorporating viewpoints from different learners (Falchikov, 2003; Gielen, Dochy, Onghena, Janssens & Decuyper, 2007). PA thus enhances learning in two ways: for the assessee whose work is assessed, because he/she receives direct feedback on how to improve; and for the assessor – by evaluating a peer's work, (s)he becomes more aware of his or her own strengths and weaknesses by comparing the peer's work against an internal representation of their own work (Nicol & Macfarlane-Dick, 2006; Panadero, Jonsson & Strijbos, 2016). In other words, PA enhances self-assessment capability (Reinholz, 2015) and students' self-regulated learning, with peers acting as co-regulators of their peers (Panadero et al., 2016). Having experienced both the assessee and assessor roles is important if students are to develop the ability to evaluate and improve their own work and acquire the skills needed for life beyond the university (Topping 1998). The studied PA learning and assessment environments in this dissertation are therefore both formative and sustainable in nature.

Nonetheless, PA is a difficult and intensive practice to implement. For example, it involves multiple social and human factors that need to be taken into account because peer assessment does not happen in isolation; it produces thoughts, actions, and emotions as a consequence of the interaction of assessees and assessors, which can have an impact on the quality of the PA process (Panadero, 2016), especially in face-to-face classroom contexts (Latané, 1981; Pope, 2005). Furthermore, to formulate and articulate judgements on a peer's work, students need to be able to offer high-quality feedback, and therefore need to develop assessor skills. Previous research has indicated that students require practice, training (Sluijsmans, 2002; Strijbos & Sluijsmans, 2010), and guidance (Hovardas, Tsivitanidou & Zacharia, 2014; Panadero, Romero & Strijbos, 2013). The realization of valuable PA practices within daily classroom practices can thus be

impeded by several factors, which will be discussed in the following sections.

A conceptual framework for facilitating and studying interactive peer feedback dialogue practices

As described above, PA practices are seen as a valuable method to enhance formative and sustainable assessment practices. However, the operationalization of dialogic feedback environments in which learners make sense of information from various sources is a challenging task. In their conceptual framework, called the feedback triangle, Yang and Carless (2013) identify three dimensions that are central for the enhancement of valuable feedback dialogue practices, but in which barriers and challenges to this enhancement can arise. These dimensions are: a), a *social-affective dimension*, which refers to the fact that feedback is a social practice in which the management of relationships represents a source of emotions influencing learners' ways of learning; b), a *cognitive dimension*, representing the content of feedback delivered, which is most central to student learning; and c), a *structural dimension*, which refers to the timing, sequencing, and modes of feedback, allied to resources for generating and providing feedback. In this dissertation, we applied an adapted version of the above-described framework for facilitating and studying dialogic PA processes; that is, PA practices with the crucial inclusion of a PF component. Three adapted dimensions can be discerned in the peer feedback triangle framework (see Fig. 1): a). the *social and interpersonal nature* of PA – conveying the social-affective dimension; b), the *quality* of the PF – conveying the cognitive dimension; and c), the *organization and management* of PA practices – conveying the structural dimension. Because these peer feedback processes are a specific operationalization of dialogic feedback processes, these dimensions are closely linked to the more broadly formulated dimensions of the above-described feedback triangle, which will be discussed in the following sections.

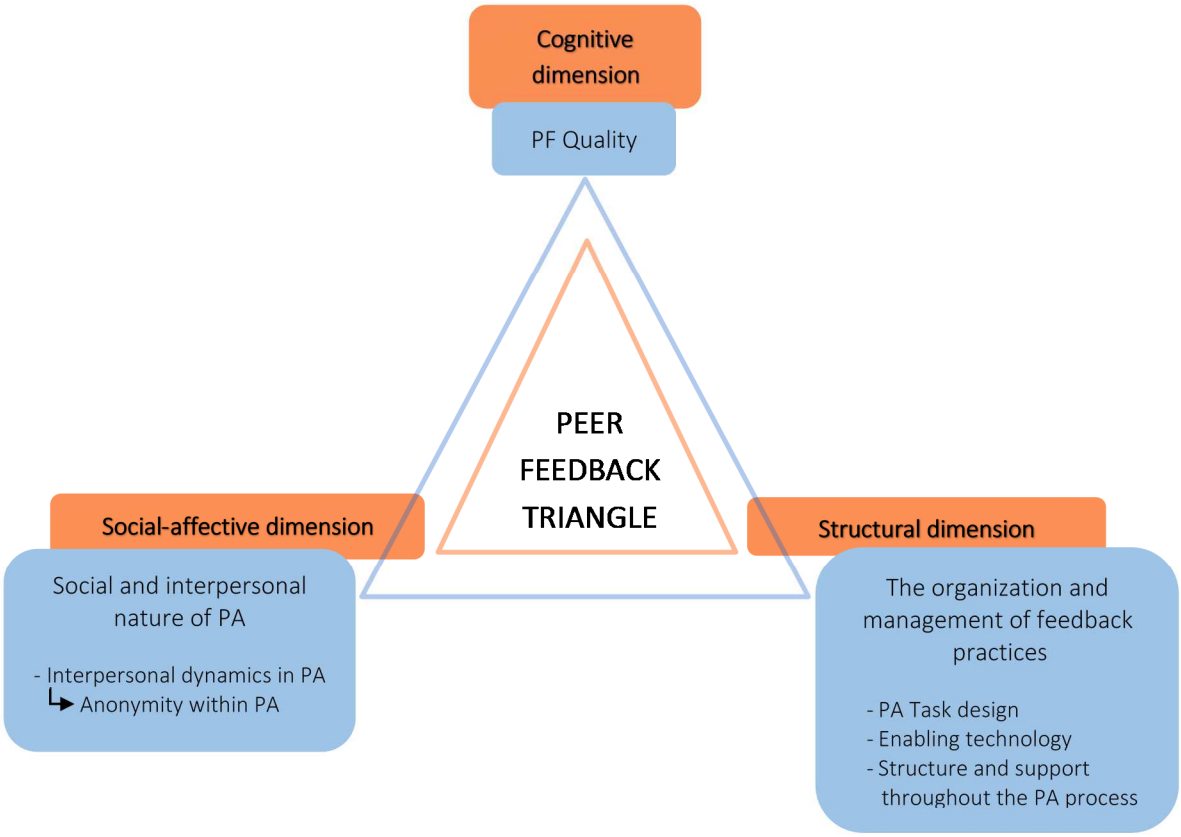


Figure 1. Peer feedback triangle

Social-affective dimension:
Social and interpersonal nature of PA

Within feedback contexts, emotions may help students to recover from negative evaluations, provide protection against rejection, make sense of feedback, and promote cooperative relationships with teaching staff and peers (Rowe, Fitness & Wood, 2014). As explored by Rowe (2017), cognitive appraisal theory offers explanations that can help to elucidate the reasons for and timing of specific emotional reactions in response to different teaching and learning contexts, such as (peer) feedback. Cognitive appraisal models propose that emotions arise as a function of people's appraisals, or interpretations of particular situations/events. This dissertation similarly adopts a cognitive approach, conceptualizing emotions as responses to appraisals of situational events or concerns. Current research paints a worrying picture of what can happen when particular human and social conditions are not taken into account when implementing assessment in education. Human conditions refer to "how individuals understand, respond to, and interpret assessment" (Harris & Brown, 2016, p. 2.). When studying the human condition, it is important to consider the beliefs, attitudes, perceptions, and/or conceptions of assessment that diverse actors hold. Additionally, one must take into account students' experiences, responses to, behaviors around, and emotions towards assessment. Social conditions refer to "how assessment is experienced in group settings [and] the interplay between the experiences of the individual and collectives to which these individuals belong" (Harris & Brown, 2016, p. 3.).

This dissertation focuses on the human and social conditions that can stimulate and/or impede the enactment of valuable PA practices by exploring students' and teachers' perceptions towards possible influencing variables. This research is needed because how to establish this necessary environment is seldom explained in depth alongside the peer assessment templates that teachers are currently are encouraged to use (Hattie, 2016). This dissertation primarily focuses on perceptions because PA is always a social experience; it does not happen in isolation, given that students define and practice assessment from their own point of view and in relation to others (e.g., teachers and peers) (McInerney, Brown & Liem, 2009). "Perceptions" is used as an umbrella term in this dissertation; it involves conceptions, attitudes, and beliefs about assessment, which are all shaped by mental interpretations about perceived information and stimuli (Bonner, 2016).

Interpersonal variables within PA

When involving students in PA, interpersonal dynamics within assessment (e.g., trust in the other, psychological safety, reciprocity effects due to friendship between peers) have to be taken into account when students assess others in front of their classmate(s). Dealing with the effects of

these interpersonal processes requires assessment literacy/capabilities for both students and teachers (Barron, 2003; Cartney, 2010; Panadero et al., 2013; van Gennip, Segers & Tillema, 2009). Well-implemented PA should decrease negative social problems, assure accuracy, and lead to positive learning outcomes (Harris & Brown, 2016; Panadero & Brown, 2017; Stepanyan, Mather, Hamilton-Jones & Lusuardi, 2009; Topping, 2010). According to the social impact theory of Latané (1981), the effects of interpersonal variables can be even stronger in face-to-face classroom settings, in that social influences, such as experiencing peer pressure, increases when others are close by. This is confirmed by several researchers, who have found that students in face-to-face classroom settings often do not feel comfortable and can experience stress when publicly evaluating their peers (Pope, 2005; Stepanyan et al., 2009).

The "problem" with emotions in PA does not lie with the individual but rather the pedagogy; that is, in the enactment of the PA practice. This dissertation will now look at how PA environments can be adequately adapted. Studying how students' perceptions of how interpersonal variables influence how they perceive the educational value of PA, both broadly and also within specific face-to-face PA settings, can broaden our insight and inform practice as concerns how to overcome this barrier. To find patterns in how students react to and interact with the peer assessment process, six interpersonal variables frequently referred to in the research will examined. Recognition of the presence of interpersonal processes in PA is thus important because the pressure students experience in the process in turn may directly impact how they view the value of PA (Li, 2016). A total of six variables, chosen because they are the most relevant when it comes to their possible effects on PA, are as follows:

(1) *Friendship marking and peer pressure*, due to friendship bonds, has been mentioned as a source of potential scoring/feedback bias in PA. However, only a small number of studies have directly addressed this issue (Panadero et al., 2013; Vanderhoven, Raes, Montrieux, Rotsaert & Schellens, 2015).

(2) *Fear of disapproval* refers to assessors' fear of negative comments from the assessee if they give them a low score or critical feedback (recrimination) (Cartney, 2010).

(3) *Psychological safety* refers to a situation in which students have a shared belief about taking interpersonal risks in a group. People who feel psychologically safe tend to perceive differences in opinions as opportunities rather than conflicts (Nicol, 2010; Yu & Sung, 2015).

(4) *Value congruency* refers to the importance of unanimity among peers in both the goals and criteria of the PA activity (Cheng & Tsai, 2012). This congruency is seen as important because previous studies have indicated that low diversity on common goals contributes to a better performance of the assessment task (Ching & Hsu, 2011).

(5) *Trust in themselves as an assessor* refers to the assessors' beliefs about their skills when assessing a peer (van Gennip, Segers & Tillema, 2010). This interpersonal factor is concerned with learners' perception of a shared responsibility for the process of evaluating peers (Cheng & Tsai, 2012).

(6) *Trust in the other as an assessor* refers to confidence in the reliability and validity of the assessment and feedback received from a peer. Students will only act on the basis of trustworthy information; if they believe that comments are capricious, they will not act on the basis of them (Carless, 2013).

Anonymity within PA

When attempting to create PA environments in which the interplay of interpersonal variables does not negatively affect the assessment outcomes, it has often been stated that offering anonymity to assessors is desirable (Ballantyne, Hughes & Mylonas, 2002; Vickerman, 2009) or should at least be studied (Howard, Barrett & Frick, 2010; Topping, 1998). The theoretical foundations with regard to the possible impact of anonymity imply that students will enact different feedback behavior, depending on whether their identity as assessors is revealed (Yu & Sung, 2015). It is this approach that the current dissertation adopts; Yu & Sung (2015) provide two perspectives from social psychology that help to support and contextualize this adoption.

Social Identity Theory (SIT) suggests that next to his/her unique personal identity as an individual, a person also forms a social identity according to the groups with which he or she affiliates (Pearce, 2013). SIT has been used to explain and predict certain personal behaviors on the basis of, among other factors, interpersonal relationships in group situations (Hogg et al., 2006). More specifically, SIT proposes that a person with a more well-received social identity and greater charisma would be perceived as the most reliable source of normative information (e.g., peer feedback), and thus have more influence over the behavior of other group members (Hogg, van Knippenberg & Rast, 2012). Additionally, adolescents and young adults are known to be particularly influenced by the views of their peers, compared to younger children (Brown, 2004). In consideration of these effects, it has been suggested that offering anonymity to participants during group interactions could foster higher participation and more balanced engagement among individuals (Chester & Gwynne, 2006; Hosack, 2004). Additionally, anonymity is believed to offer a sense of psychological safety (Miyazoe & Anderson, 2011; Yu & Liu, 2009), which is defined as a shared belief denoting one's emotional ability to take an interpersonal risk without fearing negative consequences with regard to one's well-being, self-image, and status (Kahn, 1990; Zhang, Fang, Wei & Chen, 2010). In general, individuals who feel psychologically safe are more likely to perceive differences in opinions as opportunities rather than conflicts, and to provide candid and critical peer feedback that can lead to higher-quality learning outcomes (Lu &

Bol, 2007; van Gennip et al., 2009). In addition, the negative influences that self-consciousness can exert on the assessor in peer assessment activities can also be relieved (Roberts & Rajah-Kanagasabai, 2013; Zhang et al., 2010). With these liberating effects, anonymity can help relax the social customs and conventional roles that are usually expected of students (Miyazoe & Anderson, 2011), and is thus preferable when participants know each other (e.g., Hosack, 2004). In the case of peer feedback, anonymity should allow students to express feedback that may differ from a prevailing group norm or the views of a dominant individual.

Recent studies that examine this topic are scarce and highly context-dependent, given that anonymity can be operationalized in many formats. For example, in a series of studies by Yu and colleagues (e.g., Sung, Chang, Chang & Yu, 2010), no significant difference was found on the actual interaction behavior of participants in non-anonymous and anonymous online PA conditions. Differently, Howard et al. (2010) found in an online asynchronous PA setting that students who acting anonymously were approximately five times more likely to provide critical feedback than those whose identities were known to their recipients. Other studies have indicated that providing anonymity for assessors can help to relieve the interpersonal burden on students (Yu & Liu, 2009). Cheng and Tsai (2012) found that anonymity is preferred in order to avoid the pressure of friendships. The relationship between the attributed importance of anonymity and the quality of feedback within anonymous settings has, to date, not been sufficiently explored. Furthermore, there is a gap in research on PA anonymity within real face-to-face PA settings because recent studies have mostly focused on computer-mediated discussion or online PA settings (Ainsworth et al., 2011). Therefore, the intervention studies in this dissertation will focus on offering anonymity to the assessor when providing feedback to his/her peers in face-to-face settings.

Naturally, challenges such as the creation of anonymity within face-to-face PA that arise in this social-affective dimension of the PF triangle are closely linked and have repercussions for the challenges in the structural dimension, because resources have to be found to establish anonymity within the face-to-face context. This interlinkage between the two dimensions will be considered when discussing the structural dimension.

Cognitive dimension:
The quality of PF

A second dimension that is central in the establishment of valuable PA practices is the quality of peer feedback because not all students will be equally proficient peer assessors from the start. The quality of PF is important because it is the basis for PA and provides a platform for engaging students in an interactive and elaborative feedback discourse, as well as in taking ownership of their learning (Hattie & Gan, 2011). Additionally, PF has the advantage of bringing students to a situation in which they "are on the same wavelength," making PF more understandable and useful for them (Topping, 2003). Furthermore, when writing feedback, students have more opportunities to engage in important cognitive activities such as critical thinking (i.e., deciding what contributes a good piece of work), planning, monitoring, and regulation (Lin, Liu & Yuan, 2001).

Peer feedback quality can be approached in two ways: (a), in terms of accuracy, consistency across assessors, and/or concordance with teacher feedback (Van Steendam, Rijlaarsdam, Sercu & Van den Bergh, 2010), or (b), in terms of content and/or style characteristics (Gielen, Peeters, Dochy, Onghena & Struyven, 2010). The first approach focuses on the numbers of errors and/or holistic scores for correctness of peer comments. This definition originates from the summative view on PA, where scoring validity and reliability are the primary goals (Gielen et al., 2010). In our opinion, this view is problematic because peers are inevitably novices and not experts, unless PA is transformed into cross-level peer tutoring. The problem for some teachers/researchers is that students, as novices and learners, may not be sufficiently competent in a field to make an accurate estimation of another's quality of work. However, empirical research shows that students can be reliable sources under appropriate conditions (Falchikov & Goldfinch, 2000; Topping, 2003), such as with the use of rubrics and involving students in discussions about the criteria (Panadero et al., 2013). In a recent survey of teachers' reasons for PA use, the results clearly demonstrate that teachers are concerned about the accuracy of PA, although it was not a significant predictor of the self-reported use of PA (Panadero & Brown, 2017). Because this dissertation does not focus on PA as a marking activity, the broad term "accuracy" will be used when exploring this factor in relation to the social interactive nature of PA because it refers to the overall quality, in conjunction with the reliability and validity indicators. The second approach with regard to PF quality, which is supported in this dissertation, defines PF quality in terms of content characteristics. Gielen and colleagues state that "the advantage of this approach is that such characteristics are not domain- or task-specific, thus teaching students to focus on content and style characteristics results in a generic skill transferable to other settings" (Gielen et al., 2010, p. 306). In other words, this approach focuses on the development of students' evaluative expertise, preferably beyond the

immediate task (cf. Boud & Soler, 2015; Carless, Salter, Yang & Lam, 2011). Nicol, Thomson and Breslin (2014) add that to fully realize the benefits of peer feedback, students must produce a written explanation for their evaluative judgements; producing explanations is a constructive learning activity. Building on recent work by Gielen and De Wever (2015), in this dissertation, we focus on the content of PF messages. Previous research indicates that qualitative feedback should provide two types of information: verifications and elaborations (Narciss, 2008). Verification refers to "a dichotomous judgment to indicate that a response is right or wrong," while elaboration means "relevant information to help the learner in error correction" (Hattie & Gan, 2011, p. 253). These types of information are thus seen as the structural components of feedback, because students require feedback that tells them not only if they have performed the task correctly, but also why and what they should do to improve their work (e.g., Prins, Sluijsmans, Kirschner & Strijbos, 2010). Therefore, offering elaborations that justify the verification (e.g., correct vs. incorrect) is presumed to be beneficial for students' learning. As a consequence, a balanced proportion of verifications and elaborations is more valuable than providing verifications alone (Gielen & De Wever, 2015). Gielen and De Wever (2015) explored PF quality with higher education students in an asynchronous (i.e., non-immediate PF) wiki environment. Because of the rapid development of computer technology in educational settings (Yang & Tsai, 2010), most recent research has focused on online asynchronous PA, partly neglecting the fundamental difference with face-to-face synchronous PA settings (van Popta, Kral, Camp, Martens & Simons, 2017). For this reason and in relation to the challenges for research and practice that were formulated when describing the social-affective dimension of PA, this dissertation will focus on PF quality in synchronous face-to-face PA settings in which immediate PF is given.

Structural dimension: Organization and management of PA practices

This third dimension refers to the organization, timing, sequencing, and modes of feedback, allied to non-disciplinary resources for generating and providing feedback (Yang & Carless, 2013). The structural constraints that can arise are multimode, due to the enormous diversity in PA types (e.g., process-related, product-related, group-based, individual-based, same-year, and cross-year), the classroom setting (large vs. small class size), and timing (short- vs. long-time commitment). These realities can hinder students from engaging in dialogic PA processes (Yang & Carless, 2013). Therefore, it is important for educational researchers to inform teachers how to mitigate these structural barriers through (re)engineering PA practices. In this dissertation, we will focus on three components that are frequently focused on in the (re)design of these practices: a), PA task design; b), enabling technology; and c), structure and support.

PA task design

PA task design refers to the design decisions that have to be made when implementing a PA activity in class. It is widely acknowledged that alignment between instruction and assessment is necessary in order to meet the goals of education (Biggs, 1996). Constructive alignment should thus also be ensured in dialogic PA practices, where students go through a cyclical process of performing a task, assessing task performance, identifying points of improvement, and planning future tasks (Van Merriënboer & Sluijsmans, 2009). This ensures that students receive relevant feedback that can be used as "feed-forward" into future work (Liu & Carless, 2006). The specific performance task should preferably resemble real-life professional tasks (Dochy, Segers & Sluijsmans, 1999). This does not require assessment to take place in the workplace or in a genuine social context, but assessment tasks must resemble professional tasks to some degree (Gulikers, Kester, Kirschner & Bastiaens, 2008). Furthermore, the nature of the (peer) assessment task should impact on prospects for the development of evaluative expertise (Boud & Soler, 2015; Carless, 2015). An authentic (peer) assessment task allows students to see the coherence between schoolwork and work awaiting them "in the real world" (Wu, Heng & Wang, 2015). In this dissertation, we chose to let student assess each other on oral group presentations (intergroup peer assessment; Strijbos, 2016) "in which work is openly evident to peers, which provides opportunities for student appreciation of quality and associated development of evaluative expertise" (Carless, 2015, p. 966).

Enabling technology

In their description of the structural dimension of the feedback triangle, Yang and Carless (2013) refer to the challenge of finding workload-efficient means for involving students in classroom activities that develop students' evaluative capabilities in tandem with feedback from peers and tutors. Likewise, in order to explore the role of anonymity within face-to-face synchronous intergroup PA settings (see the *social-affective* dimension), the challenge of both allowing students to provide peer feedback and capturing this data and sending it out to assessees within a reasonable time demands new solutions. In recent years, educational researchers have explored the use of clickers (also known as polling devices, classroom response technology, or an electronic voting system) for PA scoring purposes (Raes, Vanderhoven & Schellens, 2013; Vanderhoven et al., 2015). However, with these systems, students lack the possibility of giving peer feedback – the essential learning component in PA. This lack of immediate – whether digitally or not – written peer feedback hinders the creation of a rich feedback and learning environment, where assessment and feedback practices are not viewed as "added on" activities (Havnes, Smith, Dysthe & Ludvigsen, 2012). As stated by Nicol (2010), it would add more "pedagogical power" to a

learning environment were it to provide both immediate (peer) assessment and feedback, which provides information on how to improve future performance and learning. The prevalence of portable technology, also known as Mobile Response Technology (MRT), might offer a solution to this. Because these new technologies can capture and easily share both student and teacher input, they could also have the potential to facilitate (peer) assessment and feedback processes (Magaña & Marzano, 2014). Furthermore, this approach eliminates the high cost of classroom technology hardware (Güler, 2016), at least when students are able and/or allowed to bring their own device and if free web-based services are used. MRT operates via a free web-based application on web-enabled devices via wi-fi or mobile Internet (i.e., smartphones, tablets, laptops, and desktops) whereby students can give scores and written feedback to each other. Frequently used software applications are PollEveryWhere, Mentimeter, Socrative, and WhatsApp (Güler, 2016). The benefits of MRT include having students see the distribution of class responses when (non-) anonymously projected live in real-time in the classroom, and that mobile devices are suitable for sending responses to open-ended questions by instant messaging (Stowell, 2015). In this dissertation, the application Socrative (ww.socrative.com) has been chosen to investigate the use of MRT for PA purposes.

Structure and support throughout the PA process

A third component of the structural dimension concerns the structure and support initiatives within PA processes. PA is often described as a complex collaborative learning task that requires high-level cognitive processing (e.g., Kollar & Fischer, 2010). Therefore, any approach to help students to provide better PF to their classmates is expected to have an impact on the implementation of PA and, finally, on learning. More research is necessary examining the type of structure and support that peer assessors need in order to produce high-quality feedback messages (Hovardas et al., 2014). Three initiatives concerning structure and/or support in PA will be explored in this dissertation.

A filter-out scaffold for the assessees

Previous research has explored two initiatives to support students in providing high-quality PF. First, offering guiding questions/guidelines on what good PF quality constitutes (Reinholz, 2015). The logic behind it is that by offering such questions, the students will reflect more about the PA exercise, which thus becomes a more metacognitive activity. This type of question can be used to help PA assessors in producing the feedback and/or by assessees to better understand the feedback received. There are multiple examples of this type of intervention, such as work by Gielen and De Wever (2012, 2015), who structured the PF process by providing a template for assessors. Another suggested approach is to help assessees filter out the useful feedback they

receive (Tsivitanidou & Constantinou, 2016). The hypothesis behind this is that when assessees actively process the PF they receive, they will become better assessors in a subsequent task, which means that they will produce better quality PF. This type of scaffold will be studied in the second intervention study in this dissertation.

Anonymity used as a scaffold to create a safe learning environment

As already mentioned in the social-affective dimension, the existing research suggests that anonymity is one of several factors that encourage student participation (Ballantyne et al., 2002). Additionally, due to its role in diminishing reciprocity effects, anonymity might result in a fairer assessment (Freeman & McKenzie, 2000). In a recent review study, Panadero (2016) correctly pointed out tension between implementing anonymous PA and the formative use of PA; anonymity might lessen the impact of interpersonal processes, while simultaneously hindering the creation of a rich and interactive feedback environment (cf. dialogic PF processes as defined by Carless). When implementing PA, teachers are thus challenged to strike a balance between the creation of a safe learning environment provided through anonymity and the creation of a rich PF setting (Panadero, 2016). To date, no clear guidelines are available for teachers to cope with this tension. Building on the suggestions of both Howard et al. (2010) and Vanderhoven et al. (2015), who state that anonymity could function as a temporary scaffold to generate critical feedback. This hypothesis is explored in the third intervention study in this dissertation by the creation of a face-to-face PA activity in which higher education students experience the transition from an anonymous to a non-anonymous setting.

Students' preferred type of teacher assessment and feedback

Furthermore, there is a paucity of research on PA regarding the factors that operate to constrain or enable teachers' beliefs in the educational value of PA practices (Parr & Timperley, 2016). Harris and Brown (2013) suggest that the successful implementation of PA practices depends on the teacher's ability to adequately prepare students for this activity. Hovardas, Tsivitanidou and Zacharia (2014) point out that guidance and support throughout the whole process might be crucial. The tension for the teacher in this regard lies in finding the balance between on the one hand leaving the expert role as a teacher-assessor, and on the other hand passing the process completely into the hands of the students. Previous research indicates that students used to a teacher-led assessment can become frustrated (Gielen et al., 2007) or experience discomfort (Vanderhoven et al., 2015) when the teacher's assessment and feedback opportunities are completely replaced by peer assessment and feedback. Because of the specific focus on the role of anonymity within PA (cf. social-affective dimension), in the first intervention study in this dissertation, students' perceptions when teacher's assessment and feedback is given

anonymously and thus intertwined with the assessment and feedback of peers, compared to a setting in which the teacher's assessment and feedback is given non-anonymously, will be investigated.

Research design and overview of the dissertation:

Research aim & objectives

Building on the shortcomings and challenges identified within the three dimensions of the PF triangle, this dissertation aims to examine the interpersonal nature of face-to-face PA processes and find answers concerning how to create a safe and useful PA environment in which students can practice their evaluative judgment skills. These goals will be operationalized by examining both students' and teachers' perceptions towards interpersonal processes and analyzing the quality of students' PF messages, when involved in anonymous face-to-face intergroup PA settings. These research aims are considered particularly in the context of secondary and higher education settings. The following research objectives are put forth:

Research objective 1 (RO1):	To determine the current use of PA practices in secondary education in Flanders and investigate the relationship between participants' PA conceptions and its social nature.	*Social-affective + structural dimension*
Research objective 2 (RO2)*:*	To investigate students' perceptions of interpersonal variables and the importance attributed to anonymity, when involved in a face-to-face PA facilitated by MRT.	*Social-affective + structural dimension*
Research objective 3 (RO3):	To examine students' PF quality within face-to-face PA settings.	*Cognitive dimension*

These research objectives are considered more deeply in chapters two to six. Table 1 indicates which research objectives are discussed in which chapter.

Table 1 *Overview*

RO	Dimension	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
RO1	Social-affective + structural	*	*			
R02	Social-affective + structural			*	*	*
RO3	Cognitive				*	*

Design of the studies and overview of the dissertation

The studies considered in this dissertation can be typified as descriptive and quasi-experimental research. Descriptive studies use methods to describe and interpret what exists in the present, whereas quasi-experimental studies investigate intervention effects in classroom settings. The included studies represent quantitative survey studies, whereby quantifiable data is collected to be statistically analyzed in an objective manner (Creswell, 2008). Next to this, mixed-methods approaches are used, in which quantitative measures are combined with participants' interpretations and personal experiences situated and embedded in context. For the content analysis of students' peer feedback messages, data transformation is applied, whereby qualitative data are converted into quantities that can be analyzed statistically (Teddlie & Tashakkori, 2010). Data from the quasi-experimental studies were collected in both secondary (11th grade) and higher education settings. In both settings, students were asked to perform an oral group presentation or workshop. Subsequently, they were asked to assess their peers by means of rubric scores accompanied by peer feedback, via MRT. Table 2 offers an overview of the different chapters, in which more details of the research objectives, research questions, methodology, research design and sample, data collection, and data analysis techniques are shown. The following paragraphs will discuss the outline of the seven chapters.

Chapter 1 is the general introduction of the present dissertation, wherein the conceptual framework and the resulting research objectives are outlined. Furthermore, an overview of the design, methodology and studies included in the dissertation is provided.

In line with research objective 1 (RO1), chapter 2 – titled "*How do students perceive the educational value of Peer Assessment in relation to its social nature? A survey study in Flanders"* – aims (1), to map current use of PA practices in Flemish secondary education, as expressed by students and (2), to explore students' perceptions of PA interpersonal variables, anonymity, and accuracy and their relationship with beliefs on the educational value of PA. In a quantitative cross-sectional survey study, data were collected from 3680 students at all levels and educational types in the Flemish secondary education. For this purpose, Structural Equation Modelling (SEM) was used to explore what variables predict students' PA conceptions. In addition, tests of mean latent differences were performed to explore between educational levels, PA experience, and gender. This chapter is published in the journal *Studies in Educational Evaluation*.

Similarly to chapter 2, in chapter 3, teachers' awareness of students' concerns on these interpersonal processes are studied. Again, in a quantitative cross-sectional survey study, data were collected from 225 teachers from all levels and educational types in Flemish secondary education. This chapter is under review for publication in the journal *Studies in Educational Evaluation.*

Chapter 4 explores the use of MRT as a facilitator of immediate anonymous PA and PF in a face-to-face classroom setting in higher education. A second aim is to investigate how the interpersonal processes are related to students' conceptions of PA and their perceived learning in PA (RO2). Third, in search of the complementarity of peer feedback and teacher feedback in PA settings and taking into account the possible identity revelation modes that MRT offers, students' preferences towards (non-) anonymous teacher assessment and feedback were studied. In this study, 39 university students received a group assignment to prepare and present a workshop on a provided topic. Teams consisted mostly of two or three students. A within-subject design manipulation was set up, in which students experienced (from an assessor's perspective) two types of teacher assessment and feedback (i.e., *anonymous teacher assessment and feedback* vs. *non-anonymous teacher assessment and feedback)* during two consecutive sessions. A mixed-method approach was used to obtain a complete picture of the intervention. This chapter is based on an article that was submitted for publication in the *Journal of Educational Computing Research.*

Chapter 5 explores the effects of PA practice on the PF quality of 11th grade secondary education students (n=36) (RO3). Anonymous assessors gave an immediate PF using MRT during multiple feedback occasions, spread over two semesters. The design was quasi-experimental (experimental vs. control condition), in which students in one condition received a scaffold to filter out relevant information they received. It was explored whether this filter-out scaffold would influence PF quality in subsequent tasks in which they were assessors. To measure the evolution of PF quality over time, the feedback content was analyzed at the individual level on three occasions. This resulted in a database of 1,561 feedback segments. The qualitative content data were treated quantitatively, and repeated measures ANOVAs were performed for all content categories. This chapter is based on an article published in the *European Journal of Psychology of Education.*

Next, chapter 6 investigates the effects of fading anonymity on PF quality (RO3) while exploring university students' perceptions about the process (RO2). In this study, anonymity is used as an instructional scaffold for the enhancement of a safe PA environment in which students learn to provide non-anonymous high-quality feedback. Forty-six bachelor students in Educational Studies participated in multiple PA cycles in which groups of students assessed each other's work. Content analysis was performed on 4,390 feedback segments. Repeated measures ANOVAs were used to analyze students' evolution in PF quality, as well as students' perceptions about the process. This chapter is based on an article that is currently under minor revision to be considered for publication in a special issue of the *European Journal of Psychology of Education.*

Finally, chapter 7 provides a general discussion and conclusion of the dissertation. It synthesizes and discusses the main findings of the preceding chapters in relation to the structural,

social-affective, and cognitive dimensions of the PF triangle. Additionally, limitations and possible directions for future research are proposed. The chapter concludes with the dissertation's contributions to and implications for research, practice, and policy.

Table 2 *Research objectives, research questions, methodology, research design and sample, data collection and data analysis techniques.*

Chapter	Study	Research objective	Research questions	Research design and sample	Data-collection	Data analysis technique
1			General introduction			
2	1	**RO1**	**RQ1** - Do Flemish secondary education students report having experienced PA, with what frequency, and in what format? **RQ2** – What interpersonal variables predict students' PA conceptions? **RQ3** – What differences exist between gender, educational levels and students with differing PA experience?	Cross-sectional survey in Secondary Education (*n = 3680)*	Self-report method	EFA CFA Multiple Regression SEM
3	2	**RO1**	**RQ1** - Do Flemish secondary education teachers report having experienced PA, with what frequency, and in what format? **RQ2a** – Are teachers aware about students' perceptions on interpersonal processes within PA? **RQ2b** – What variables predict the educational value teachers attribute to PA?	Cross-sectional survey in Secondary Education (*n = 225)*	Self-report method	EFA CFA Multiple Regression SEM
4	3	**RO2**	**RQ1** - How do students experience anonymous PA with the use of MRT? **RQ2** - What is the impact of interpersonal variables and conceptions of peer assessment in relation with perceived learning in PA? **RQ3** - What type of teacher assessment and feedback do students prefer within this PA setting?	Intervention with pretest– posttest design (*n = 39*)	Questionnaires (closed and open-ended questions)	Descriptive analysis Content Analysis Multiple linear Regression
5	4	**RO 2 & 3**	**RQ 1** - What is the evolution in PF quality over time when students practice PA several times? **RQ2** - What is the impact of helping assessees to filter out the feedback they receive on their own PF skills as assessors? **RQ3** - Do the perceived PA skills change over time? Are they related to the actual change in the PF quality? **RQ4** - Did the students perceive the PF as useful?	Intervention with pretest– between - posttest design (*n = 36*) Repeated-measures design	Questionnaires (closed and open-ended questions) Peer Feedback Segments (*n* = 1561)	Descriptive analysis Content Analysis Repeated measures ANOVA
6	5	**RO 2 & 3**	**RQ1** - How does PF quality change over time when students consecutively practice PA in an anonymous and non-anonymous setting? **RQ2** - Do students' perceived PF skills change over time in a PA setting with a transition from anonymous to non-anonymous? **RQ3** - How does the transition from an anonymous to a non-anonymous PA affect students' perceptions regarding a) importance of anonymity, b) their perceptions towards interpersonal variables and c) their general conceptions towards PA?	Intervention with pretest– between - posttest design (*n = 46*) Repeated-measures design	Questionnaires (closed and open-ended questions) Peer Feedback Segments (*n* = 4390)	Descriptive analysis Content Analysis Repeated measures ANOVA
7			General discussion and conclusion			

References

Ainsworth, S., Gelmini-Hornsby, G., Threapleton, K., Crook, C., O'Malley, C., & Buda, M. (2011). Anonymity in classroom voting and debating. *Learning and Instruction*, *21*(3), 365–378. doi: 10.1016/j.learninstruc.2010.05.001

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing Procedures for Implementing Peer Assessment in Large Classes Using an Action Research Process. *Assessment & Evaluation in Higher Education*, *27*(5), 427–441. doi: 10.1080/0260293022000009302

Bandura, A. (1997). Self-efficacy: The Exercise of Control. *Encyclopedia of Human Behavior*, *4*, 71–81. doi:10.1002/9780470479216.corpsy0836

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, *12*, 307–359. doi: 10.1207/S15327809jls1203_1

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347–364. doi: 10.1007/BF00138871

Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising New Modes of Assessment: In Search of Qualities and Standards* (pp. 13–37). Dordrecht, The Netherlands: Kluwer Academic.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. doi: 10.1080/0969595980050102

Bonner, S. (2016). Teachers' Perceptions about Assessment: Competing Narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 21–39). New York, NY: Routledge.

Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151–167. doi.org/10.1080/713695728

Boud, D. (2007). Reframing assessment as if learning were important. *Rethinking Assessment in Higher Education: Learning for the Longer Term*, 14–25.

Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 1–14. doi: 10.1080/02602938.2015.1018133

Brown, B. (2004). *Adolescents' relationships with peers. In R. Lerner & L. Steinberg (Eds.),. Handbook of adolescent psychology*. Retrieved from https://books.google.be/books?hl=nl&lr=&id=_wXasY1HyAYC&oi=fnd&pg=PA363&dq=Brown,+B.+(2004).+Adolescents'+relationships+with+peers&ots=DS5weZ5OXt&sig=vlZ1kSZVrFyq0Li4JcpDJYGHwBY

Brown, S. (2004). Assessment for Learning. *Learning and Teaching in Higher Education*, *1*, 2004–5.

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning - a Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. doi: 10.3102/00346543065003245

Carless, D. (2013). Trust and its role in facilitating dialogic feedback. In *Feedback in Higher and Professional Education* (pp. 90–103). doi:10.4324/9780203074336

Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, *69*(6), 963–976. doi:10.1007/s10734-014-9816-z

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education.* doi:10.1080/03075071003642449

Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education*, *35*(5), 551–564. doi:10.1080/02602931003632381

Cheng, K. H., & Tsai, C. C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology*, *28*, 599–618.

Chester, A., & Gwynne, G. (2006). Online Teaching: Encouraging Collaboration through Anonymity. *Journal of Computer-Mediated Communication*, *4*(2), 0–0. doi:10.1111/j.1083-6101.1998.tb00096.x

Ching, Y.-H., & Hsu, Y.-C. (2011). Design-Grounded Assessment: A Framework and a Case Study of Web 2.0 Practices in Higher Education. *Australasian Journal of Educational Technology*, *27*(5), 781–797.

Clark, I. (2012). Formative Assessment: Assessment Is for Self-regulated Learning. *Educational Psychology Review*, *24*, 205–249. doi:10.1007/s10648-011-9191-6

Cowie, B., Moreland, J., & Otrel-Cass, K. (2013). *Expanding Notions of Assessment for Learning: Inside Science and Technology Primary Classrooms*. Springer Science & Business Media.

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research.* Upper Saddle River, NJ: Pearson/Merrill Education.

Derry, S. J. (1996). Cognitive Schema Theory in the Constructivist Debate. *Educational Psychologist*, *31*(3/4), 163–174. doi:10.1080/00461520.1996.9653264

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, *69*, 145–186. doi:10.2307/1170673

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education*, *24*, 331–350. doi: 10.1080/03075079912331379935

Falchikov, N. (2003). Involving students in assessment. *Psychology Learning & Teaching*, *3*, 102–108.

Falchikov, N., & Goldfinch, J. (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, *70*(3), 287–322. doi:10.3102/00346543070003287

Fastré, G. M. J., van der Klink, M. R., Sluijsmans, D., & van Merriënboer, J. J. G. (2013). Towards an integrated model for developing sustainable assessment skills. *Assessment & Evaluation in Higher Education*, *38*(5), 611–630. doi:10.1080/02602938.2012.674484

Freeman, M., & McKenzie, J. (2000). Self and Peer Assessment of Student Teamwork: Designing, implementing and evaluating SPARK, a confidential, web based system. In *Flexible Learning for a Flexible Society*.

Gielen, M., & De Wever, B. (2012). Peer Assessment in a Wiki: Product Improvement, Students' Learning And Perception Regarding Peer Feedback. *Procedia-Social and Behavioral Sciences*, *69*, 585–594.

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, *52*, 315–325. doi:10.1016/j.chb.2015.06.019

Gielen, S., Dochy, F., Onghena, P., Janssens, S., & Decuyper, S. (2007). A complementary role for peer feedback and staff feedback in powerful learning environments. In S. Gielen (Ed.), (pp. 157–199). Leuven: Katholieke universiteit Leuven,Faculteit Psychologie en Pedagogische Wetenschappen Centrum voor Opleidingsdidactiek.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, *20*(4), 304–315. doi:10.1016/j.learninstruc.2009.08.007

Güler, C. (2016). Use of WhatsApp in Higher Education: What's Up With Assessing Peers Anonymously? *Journal of Educational Computing Research*. doi:10.1177/0735633116667359

Gulikers, J. T. M., Kester, L., Kirschner, P. A., & Bastiaens, T. J. (2008). The effect of practical experience on perceptions of assessment authenticity, study approach, and learning outcomes. *Learning and Instruction*, *18*(2), 172–186. doi.org/10.1016/j.learninstruc.2007.02.012

Harris, L. R., & Brown, G. T. L. (2016). Volume Introduction: The human and social experience of assessment: Valuing the person and context. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook*

of Human and Social Conditions in Assessment (pp. 1–17). New York, NY: Routledge.

Hattie, J. (2016). Foreword. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. IX–XVI). New York, NY: Routledge.

Hattie, J., & Gan, M. (2011). *Instruction based on feedback*. *Handbook of research on learning and instruction*. Routledge. doi:10.4324/9780203839089.ch13

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. doi.org/10.3102/003465430298487

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, *38*, 21–27. doi:10.1016/j.stueduc.2012.04.001

Hogg, M. A., Fielding, K. S., Johnson, D., Masser, B., Russell, E., & Svensson, A. (2006). Demographic category membership and leadership in small groups: A social identity analysis. *The Leadership Quarterly*, *17*(4), 335–350. doi: 10.1016/j.leaqua.2006.04.007

Hogg, M. A., van Knippenberg, D., & Rast, D. E. (2012). The social identity theory of leadership: Theoretical origins, research findings, and conceptual developments. *European Review of Social Psychology*, *23*(1), 258–304. doi:10.1080/10463283.2012.741134

Hosack, I. (n.d.). The Effects of Anonymous Feedback on Japanese University Students' Attitudes towards Peer Review.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers and Education*, *71*, 133–152. doi:10.1016/j.compedu.2013.09.019

Howard, C. D., Barrett, A. F., & Frick, T. W. (2010). Anonymity to Promote Peer Feedback: Pre-Service Teachers' Comments in Asynchronous Computer-Mediated Communication. *Journal of Educational Computing Research*, *43*(1), 89–112. doi:10.2190/EC.43.1.f

Hunt, N., Hughes, J., & Rowe, G. (2002). Formative Automated Computer Testing (FACT). *British Journal of Educational Technology*, *33*, 525–535. doi: 10.1111/1467-8535.00289

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, *33*(4), 692–724. doi:10.2307/256287

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, *20*, 344–348. doi:j.learninstruc.2009.08.005

Latané, B. (1981). The psychology of social impact. *American Psychologist*, *36*(4), 343–356. doi:10.1037/0003-066X.36.4.343

Li, L. (2016). The role of anonymity in peer assessment. *Assessment & Evaluation in Higher Education*, *2938*(April), 1–12. doi:10.1080/02602938.2016.1174766

Lin, S. s. j., Liu, E. z. f., & Yuan, S. m. (2001). Web-based peer assessment: feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, *17*(4), 420–432. doi:10.1046/j.0266-4909.2001.00198.x

Liu, N.-F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, *11*, 279–290.

Lu, L., & Bol, R. (2007). A Comparison of Anonymous Versus Identifiable e-Peer Review on College Student Writing Performance and the Extent of Critical Feedback. *Journal of Interactive Online Learning*, *6*(2), 100–115. doi:10.1017/CBO9781107415324.004

Magaña, S., & Marzano, R. J. (2014). Using Polling Technologies to Close Feedback Gaps. *Educational Leadership*, 82–83.

McInerney, D. M., Brown, G. T. L., & Liem, A. D. (2009). *Student Perspectives on Assessment: What Students Can Tell Us about Assessment for Learning*. IAP. Retrieved from https://books.google.com/books?id=9iZy0exq7goC&pgis=1

Miyazoe, T., & Anderson, T. (2011). Anonymity in Blended Learning: Who Would You Like to Be? *Educational Technology & Society*, *14*(2), 175–187.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), Handbook of research on educational communications and technology (pp. 125-143). Mahwah, NJ: Erlbaum.

Nicol, D. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, *35*(5), 501–517. doi:10.1080/02602931003786559

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self- regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. doi:10.1080/03075070600572090

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102–122. doi:10.1080/02602938.2013.795518

Orsmond, P., Merry, S., & Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, *41*, 273–290. doi:10.1080/14703290410001733294

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 247–266). New York, NY: Routledge.

Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: positive experience predicts use. *European Journal of Psychology of Education*, 32(1), 133-156. Springer Netherlands. doi:10.1007/s10212-015-0282-5

Panadero, E., Jonsson, A., & Strijbos, J.-W. (2016). Scaffolding Self-Regulated Learning Through Self-Assessment and Peer Assessment: Guidelines for Classroom Implementation. In *Assessment for Learning: Meeting the Challenge of Implementation* (pp. 311–326). doi.org/10.1007/978-3-319-39211-0

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, *39*(4), 195–203. doi:10.1016/j.stueduc.2013.10.005

Parr, J. M., & Timperley, H. (2016). Teachers and Assessment: Enhancing Assessment Capability. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 95–108). New York, NY: Routledge.

Pearce, J. A. (2013). Using Social Identity Theory to Predict Managers' Emphases on Ethical and Legal Values in Judging Business Issues. *Journal of Business Ethics*, *112*(3), 497–514. doi:10.1007/s10551-012-1274-x

Planas Lladó, A., Soley, L. F., Fraguell Sansbelló, R. M., Pujolras, G. A., Planella, J. P., Roura-Pascual, N., … Moreno, L. M. (2014). Student perceptions of peer assessment: an interdisciplinary study. *Assessment & Evaluation in Higher Education*, *39*(5), 592–610. doi:10.1080/02602938.2013.860077

Pope, N. K. L. (2005). The impact of stress in self-and peer assessment. *Assessment & Evaluation in Higher Education*, *30*, 51–63.

Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. (2010). Formative peer assessment in a CSCL environment: a case study. *Assessment & Evaluation in Higher Education*, *30*(4), 417–444.

Raes, A., Vanderhoven, E., & Schellens, T. (2013). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Studies in Higher Education*, *0*(0), 1–16. doi.org/10.1080/03075079.2013.823930

Reinholz, D. (2015). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 1–15. doi:10.1080/02602938.2015.1008982

Roberts, L. D., & Rajah-Kanagasabai, C. J. (2013, November 8). "I'd be so much more comfortable posting anonymously": Identified versus anonymous participation in student discussion boards. *Australasian Journal of Educational Technology*. doi:10.1234/ajet.v29i5.452

Rowe, A. D. (2017). Feelings About Feedback: The Role of Emotions in Assessment for Learning (pp. 159–172). Springer Singapore. doi:10.1007/978-981-10-3045-1_11

Rowe, A. D., Fitness, J., & Wood, L. N. (2014). The role and functionality of emotions in feedback at university: a qualitative study. *The Australian Educational Researcher*, *41*(3), 283–309. doi:10.1007/s13384-013-0135-7

Sluijsmans, D. M. A. (2002). Student involvment in assessment: The training of peer assessment skills. (Unpublished doctoral dissertation). Open University of the Netherlands, Heerlen.

Stepanyan, K., Mather, R., Hamilton-Jones, C., & Lusuardi, C. (2009). Avoiding Pitfalls of Peer Assessment. *Icalt: 2009 Ieee International Conference on Advanced Learning Technologies*, 186–188.

Stowell, J. R. (2015). Use of Clickers vs. Mobile Devices for Classroom Polling. *Computers & Education*. doi:10.1016/j.compedu.2014.12.008

Strijbos, J. W., Ochoa, T. A., Sluijsmans, D. M. A., Segers, M. S. R., Tillema, H. H. (2009). Fostering interactivity through formative peer assessment in (web-based) collaborative learning environments. In *In C. Mourlas, N. Tsianos, & P. Germanakos (Eds.), Cognitive and emotional processes in web-based education: Integrating human factors and personalization.* (pp. 375–395). Hershey, PA: IGI Global.

Strijbos, J. (2016). Assessment of collaborative learning. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 302–318). New York, NY: Routledge.

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, *20*, 265–269.

Sung, Y.-T., Chang, K.-E., Chang, T.-H., & Yu, W.-C. (2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence*, *33*, 135–145. doi:10.1016/j.adolescence.2009.04.004

Teddlie, C., & Tashakkori, a. (2010). Overview of contemporary issues in mixed methods research. *Sage Handbook of Mixed Methods in Social & Behavioral Research*, 1–44. doi:10.4135/9781506335193.n1

Topping, K. J. (1998). Peer Assessment between Students in Colleges and Universities. *Review of Educational Research*, *68*(3), 249. doi:10.2307/1170598

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), (pp. 55–87). Dordrecht, The Netherlands: Kluwer Academic.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*(4), 339–343. doi:10.1016/j.learninstruc.2009.08.003

Tseng, S. C., & Tsai, C. C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education*, *49*, 1161–1174.

doi:10.1016/j.compedu.2006.01.007

Tsivitanidou, O. E., & Constantinou, C. P. (2016). A study of students' heuristics and strategy patterns in web-based reciprocal peer assessment for science learning. *The Internet and Higher Education*, *29*, 12–22. doi:10.1016/j.iheduc.2015.11.002

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, *4*, 41–54. doi:10.1016/j.edurev.2008.11.002

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, *20*, 280–290. doi.:10.1016/j.learninstruc.2009.08.010

Van Merriënboer, J. J. G., & Sluijsmans, D. M. A. (2009). Toward a synthesis of cognitive load theory, four-component instructional design, and self-directed learning. *Educational Psychology Review*, *21*(1), 55–66. doi:10.1007/s10648-008-9092-5

van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*. doi:10.1016/j.edurev.2016.10.003

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, *20*, 316–327.

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, *81*, 123–132. doi:10.1016/j.compedu.2014.10.001

Vickerman, P. (2009). Student perspectives on formative peer assessment: an attempt to deepen learning? *Assessment & Evaluation in Higher Education*, *34*, 221–230. doi:10.1080/02602930801955986

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, *37*(1), 3–14. doi:10.1016/j.stueduc.2011.03.001

Winne, P. H., & Butler, D. L. (1994). Student cognition in learning from teaching. In *International encyclopedia of education* (pp. 5738–5775).

Wu, X. V., Heng, M. A., & Wang, W. (2015). Nursing students' experiences with the use of authentic assessment rubric and case approach in the clinical laboratories. *Nurse Education Today*, *35*(4), 549–555. doi:10.1016/j.nedt.2014.12.009

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, *18*(3), 1–13. doi:10.1080/13562517.2012.719154

Yang, Y.-F., & Tsai, C.-C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction*, *20*(1), 72–83. doi:10.1016/j.learninstruc.2009.01.003

Yu, F.-Y., & Liu, Y.-H. (2009). Creating a psychologically safe online space for a student-generated questions learning activity via different identity revelation modes. *British Journal of Educational Technology*, *40*(6), 1109–1123. doi:10.1111/j.1467-8535.2008.00905.x

Yu, F.-Y., & Sung, S. (2015). A mixed methods approach to the assessor's targeting behavior during online peer assessment: effects of anonymity and underlying reasons. *Interactive Learning Environments*, 1–18. doi:10.1080/10494820.2015.1041405

Zhang, Y., Fang, Y., Wei, K.-K., & Chen, H. (2010). Exploring the role of psychological safety in promoting the intention to continue sharing knowledge in virtual communities. *International Journal of Information Management*, *30*(5), 425–436. doi:10.1016/j.ijinfomgt.2010.02.003

2 How do students perceive the educational value of PA in relation to its social nature? A survey study in Flanders

This chapter is based on:

Rotsaert, T., Panadero, E., Estrada, E., & Schellens, T. (2017). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. *Studies in Educational Evaluation*, 53, 29-40. doi: 10.1016/j.stueduc.2017.02.003

Chapter 2
How Do Students Perceive the Educational Value of Peer Assessment in Relation to Its Social Nature? A Survey Study in Flanders

Abstract

This study explores the relationship between students' perceptions of peer assessment (PA) and its social nature. A quantitative survey study (N = 3680) was conducted in secondary education in Flanders, examining the students' perceptions of PA interpersonal variables and their beliefs on the educational value of PA. The structural equation modeling (SEM) results show that the educational value students attribute to PA was positively predicted through trust in their own and their peers' evaluative capabilities, awareness of negative interpersonal processes (e.g. fear of disapproval and friendship marking), and beliefs about PA accuracy. The importance attributed to anonymity appeared to be a negative predictor of PA conceptions. Tests of mean latent differences were performed to explore the differences between educational levels, PA experience and gender.

Introduction

Current approaches to assessment have moved away from end-of-course (i.e. summative) testing to in-course (i.e. formative), improvement-oriented interactions between learners and instructors (Black & Wiliam, 1998; Havnes, Smith, Dysthe & Ludvigsen, 2012). This formative view blurs the strict distinction between instruction and assessment (Cowie, Moreland & Otrel-Cass, 2013). Formative assessment strategies stress that the active involvement of students in assessment processes is necessary. For this reason, peer assessment (PA) has been embraced as an innovative method of formative assessment and is often seen as a way to offer significant educational value for learning (e.g. Topping, 2010).

In a PA activity, a student generates feedback that might be useful to the assessee, and potentially gives the peer assessor insights as to how their own work might be improved (Dochy, Segers & Sluijsmans, 1999; Panadero, 2016; Reinholz, 2015). In this study, PA is defined as an interpersonal, collaborative learning arrangement in which students assess their fellow peers' performance by providing peer feedback (PF), which can be quantitative (e.g. grades or ratings across assessment criteria) and/or qualitative (e.g. written or oral comments) (Bolzer, Strijbos & Fischer, 2015; Topping, 2010). Despite the benefits of PA, it remains a challenging assessment

method to implement. Its social nature due to being a fundamentally interpersonal process has been relatively overlooked and has only been studied in relatively small-scale intervention studies, mainly within vocational and higher education (Panadero, 2016; van Gennip, Seger & Tillema, 2010).

Furthermore, previous research shows that teachers and students' conceptions about the purpose of assessment largely influence its implementation (Brown, Lake & Matters, 2011; Segers & Tillema, 2011). To this end, this study aims to explore the relationship between secondary student perceptions related to PA and its social nature.

Conceptions of Peer Assessment

Classroom assessment is always a social experience; it does not happen in isolation because students define and practice assessment from their own point of view, and in relation to others (e.g. teachers, peers) (Brown, McInerney & Liem, 2009). These student conceptions represent ideas, beliefs, propositions, and preferences that fundamentally describe how students experience educational matters such as assessment practices (Brown et al., 2009). In formative assessment, students' conceptions are essential because it is the students who must use the assessment to learn (Cowie, 2009). Unfortunately, the fact that student conceptions have not yet been explored in sufficient detail, even in formative assessment literature, provides a knowledge gap for the present study.

A small number of studies have been conducted on students' conceptions of assessment, indicating their significant contribution to students' learning behavior and (future) learning (e.g. Harris, Brown & Harnett, 2014; Struyven, Dochy & Janssens, 2005). In a series of studies, Brown and colleagues report on how to measure student and teachers' conceptions of the purpose of assessment (e.g. Brown & Hirschfeld, 2008). Building on Ajzen's (2005) theory of planned behavior, which suggests that personal intentions or beliefs about what others think shapes ones' behavior, Brown and colleagues argue that a student's response to assessment depends on their appreciation of the process and its aims. Additionally, drawing on Zimmermans' (2001) self-regulation theory, they state it is important for students to have a "personally meaningful purpose of assessment," as self-regulated learners often need to use feedback from educational assessment (Brown & Harris, 2011, p. 46).

Therefore, more insight into how students perceive PA is essential. A recent large-scale survey study generated a robust picture of what teachers think about the use of PA (Panadero & Brown, 2016). This study came to the conclusion that, although teachers like the instructional use of PA overall, they struggle with its inherent difficulties (e.g. peer pressure), and only use PA if previous experiences have been positive. Interestingly, primary and secondary teachers reported

higher values of PA implementation and certainty about its educational value, in contrast with the higher education teachers. Similar results were found in two previous studies that investigated how higher and secondary teachers perceived PA with smaller sample sizes (Lynch & Golen, 1992; Noonan & Duncan, 2005). Nevertheless, the field lacks knowledge about the conceptions of students as the most important actors in this fundamentally interpersonal assessment process, and will thus be explored in the present study.

Peer Assessment and Its Social Nature: Six Interpersonal Variables of Interest

The majority of research has always been aware of the inherent social and emotional nature of assessment, and this is especially the case for PA (Boud, 1995). Previous research has questioned the accuracy and/or validity and/or reliability of PA (for a detailed discussion see Panadero, Romero & Strijbos, 2013) due to the presence of possible reciprocity effects caused by interpersonal processes such as friendship marking or psychological risk (Harris & Brown, 2013; van Gennip et al., 2010; Vanderhoven et al., 2015). The limited number of studies on the social nature of PA show that students' perceptions of these interpersonal processes might be related to their conceptions of PA (Cheng & Tsai, 2012; Harris & Brown, 2013; van Gennip et al., 2010). Attention to social and human factors is thus needed because well-implemented PA should decrease negative social problems, assure accuracy, and lead to positive learning outcomes (Panadero & Brown, 2016; Topping, 2010).

Six interpersonal variables frequently referred to in research are discerned in this study (for a detailed discussion see Panadero, 2016). These six variables were chosen because they are the most relevant when it comes to their possible effects on PA, and are as follows. (1) *Friendship marking*, due to friendship bonds, has been mentioned as a source of potential scoring bias. However, only a small number of studies have directly addressed this topic (Panadero et al., 2013). Recent research on the diminishing effect of rubrics on over- and underscoring by peers in PA shows that, whilst a low or medium friendship reduces the friendship bias, for high-level friendship the rubric seems to amplify the potential friendship bias (Panadero et al., 2013). Cheng and Tsai (2012) find that anonymity is preferred for the reason of avoiding the pressure of friendships. (2) *Fear of disapproval* refers to the assessors' fear of negative comments from the assessee if they give them a low score or critical feedback (recrimination) (Cartney, 2010). To decrease this type of fear, it has been argued that anonymity might play a role. For example, in the research of Vanderhoven et al. (2015), students in an anonymous, face-to-face PA setting experienced significantly less fear of disapproval compared to students in a non-anonymous setting. (3) *Psychological safety* refers to a situation in which students have a shared belief about

taking interpersonal risks in a group. People that feel psychologically safe tend to perceive differences in opinions as opportunities rather than conflicts (Nicol, 2010; Yu & Sung, 2015). This is important as several authors state that creating a safe environment is a precondition for accurate, and thus valuable, PA activities (Harris & Brown, 2013; van Gennip, Segers & Tillema, 2009). (4) *Value congruency* refers to the importance of unanimity on both the goals and criteria of the PA activity (Cheng & Tsai, 2012). Rubrics hold the potential to augment the value congruency within a PA activity as they provide the assessment criteria in a structured format and might thus enhance the perceived fairness and comfort with PA (Panadero et al., 2013). (5) *Trust in themself as an assessor* refers to the assessors' beliefs about their skills when assessing a peer (van Gennip et al., 2010). Previous research has indicated that the more the assessor trusts himself/herself, the deeper the learning from PA becomes (Cheng & Tsai, 2012), which has the potential to be increased through intensive practice and interaction (Panadero et al., 2016). (6) *Trust in the other as an assessor* refers to the confidence in the reliability and validity of the assessment and feedback received from a peer. Students will only act on the basis of trustworthy information: if they believe that comments are capricious, they will not act on the basis of them (Carless, 2013).

Two crucial aspects in peer assessment are anonymity and accuracy. As this interplay of interpersonal variables influences the assessment outcome, it has often been stated that decreasing negative social effects via anonymity is desirable (Ballantyne, Hughes & Mylonas, 2002; Vickerman, 2009) or should at least be explored (Howard, Barrett & Frick, 2010). Topping (1998) indicates that privacy is an important structural feature of PA, in that disclosing the identity of the assessor or assessee seems to matter to students. Vanderhoven et al. (2015) find that students have more positive attitudes toward PA when anonymity for the assessor was assured, while the participating teacher reported that revealing his/her identity worked as a means to control any undesirable social effects. Yu & Sung (2015) state that anonymity might offer more psychological safety for students, but at the same it might lead to misbehavior, such positive marking toward friends. A recent survey study Panadero and Brown (2016), which explores the reasons Spanish teachers gave for using PA, reveals that the majority used an anonymous version of PA. It is worth noting, however, that this verdict was not found to be a significant determinant of the frequency of PA use, except for by university teachers. In conclusion, anonymity needs further research, especially with regard to students' conceptions of the different anonymity modes that can be manipulated in a PA setting (i.e. the anonymity of the assessor, the assessee and the teacher).

Another crucial aspect related to PA is the concern about the (perceived) validity and/or accuracy and/or reliability of students' PA. The problem for some is that students, as novices and learners, may not be sufficiently competent in a field to make an accurate estimation of another's

quality of work. Empirical research shows that students can be reliable sources under appropriate conditions (Falchikov & Goldfinch, 2000; Topping, 2003) such as the use of rubrics, involving students' in the discussion about the criteria, and/or considering the level of expertise of the students (for a detailed discussion see Panadero et al., 2013). In the aforementioned Spanish survey study on teachers' reasons for PA use, the results clearly demonstrate that teachers are concerned about the accuracy of PA, although it was not a significant predictor of the self-reported use of PA (Panadero & Brown, 2016). In the present study, the broad term "accuracy" will be used as it refers to the overall quality in conjunction with the reliability and validity indicators.

To sum up, interpersonal variables, anonymity, and accuracy play an important role in how students perceive the educational value of PA. However, our knowledge about these fundamental issues is often based on small sample studies. Therefore, there is a need for a larger study that explores these issues.

The Flemish Context and Its Assessment Practices

To understand the results of this study, it is important to get acquainted with the assessment context in Flanders (the Dutch community in Belgium). In general, there are four types of schools: general secondary education, technical secondary education, arts secondary education, and vocational secondary education. The government imposes a minimal timetable, which incorporates core-curriculum subjects depending on the educational level. Schools are distributed across educational networks: GO! (18.4%), official subsidized education (7.4%), and subsidized private education (75.5%), which functions independently of the Flemish Ministry of Education (Flemish Governement, 2015). From an international point of view, secondary education in Flanders is known for its high quality (e.g. being within the top 10 for mathematical reasoning in PISA 2012) (Department of Educational Studies–Ghent University, 2013).

At a Belgian level, national tests do exist, but these are exclusively concerned with the compulsory attainment and developmental targets of the curriculum, and have no public accountability element. The main goal of this type of national testing is to monitor and evaluate schools and/or the education system as a whole. National test results are thus used as indicators of the quality of teaching and the performance of teachers, but also to gauge the overall effectiveness of educational policies and practices. It is expected that this specific context of national testing could have an influence on students' conceptions of peer assessment.

On both a meso and micro level, for decades the Flemish government has emphasized the importance of autonomy and trust in the policy-making capacity of schools. As a part of this autonomy, teachers and teacher councils are, as a rule, solely responsible for the majority of pupil

learning and classroom assessment. All assessment practices are thus voluntary and low stakes. As a consequence, implementing formative assessment is the responsibility of individual teachers.

Regardless of the high amount of policy autonomy that has been attributed to school boards, the Flemish government has taken some initiatives to promote formative assessment in the curriculum, such as the development of the toolkit *Breed Evalueren* ("broad assessment"), which aims to support teachers in assessing Dutch competences in both primary and secondary education. PA is explicitly mentioned herein (De Backer & Philips, 2013). Furthermore, the two biggest educational organizations have outlined – each for their own organization - a clear and balanced vision on assessment, whereby both formative assessment and assessment *of* learning practices are promoted (GO!, 2012; Katholiek Onderwijs Vlaanderen, 2013).

As these recent initiatives promote formative assessment practices and give concrete suggestions on how to implement them, it could be assumed that formative assessment practices have found their way into classroom practice. Therefore, it could be expected that Flemish secondary education students would understand the educational value of assessment practices and have already had some PA experience. However, no up-to-date empirical evidence has yet confirmed the effects of the aforementioned initiatives.

Aim and Research Questions

The present study will explore Flemish secondary education students' self-reported experiences with PA, and aims to determine whether the concerns raised in the literature exist and affect students' attitudes toward PA. The research questions are:

RQ1–Do Flemish secondary education students report having experienced PA, with what frequency, and in what format?

RQ2–What variables predict students' PA conceptions?

Based on the previously described theoretical framework, it is hypothesized that (a) students' perceptions of negative interpersonal variables (e.g. friendship marking,) will negatively affect their belief in the educational value of PA, (b) positive interpersonal variables (e.g. psychological safety, trust, and value congruency) will positively affect students' PA conceptions, (c) perceived validity/accuracy will increase their belief in the educational value of PA, and (d) a perceived high importance for the use of anonymous forms of PA will decrease positive beliefs in the educational value of PA.

RQ3– What differences are there in the relations among students' perceptions of interpersonal variables, perceived accuracy, the attributed importance of anonymity and the perceived educational value of PA that exist between gender, educational levels, and students with different PA experiences?

Method

Participants

A total of 3,680 Flemish high school students participated in this study, with a subsample of 3,066 students with PA experience. In this subsample, 32.2% of the participants were in grades 7-8 (henceforth *Level 1*); 34.4% in grades 9-10 (henceforth *Level* 2), and 33.5% in grades 11-12 (henceforth *Level 3*). In terms of the demographic information, the students had an average age of 15.10 (*SD= 1.94*), ranging from 11–21 years old. The percentages of male and female students was 45.3% (N=1398) and 54.6% (N=1675), respectively. The distribution of the collected data over the four educational types was 52.5% in general secondary education, 28.7% in technical secondary education, 13.4% in vocational secondary education, and 5.4% in arts secondary education. In addition, distribution by Flemish regions can be seen in Table 1. The majority of the data was collected in East Flanders because the university that organized this survey is situated there and it was easier to arrange the data collection.

Table 1
Distribution of the sample by regions and type of school

Region	*N*	*%*
Antwerp	*386*	10.50
East Flanders	*1702*	46.30
West Flanders	*1278*	34.70
Flemish Brabant & BCR	*216*	5.90
Limburg	*98*	2.70
Type of school		
General Secondary	*1925*	52.54
Technical Secondary	*1053*	28.74
Arts Secondary	*197*	5.38
Vocational Secondary	*489*	13.35

Procedure

The survey conductors were bachelor's degree students in educational studies who were enrolling in a practice-based methodology course. In groups of four, they were asked to get permission from eight teachers to conduct a survey in their classes. In order to reduce social desirability, the pen-and-paper survey was designed to be filled in anonymously. Participants were asked to also fill in an informed consent, which contained information about the purpose, confidentiality assurances, and the possibility to withdraw. The consent forms were collected separately so as to ensure confidentiality. In line with the protocol for scientific research at the faculty overseeing this study, the consent of the parent or the guardian was restricted to a passive informed consent.

The survey conductors received detailed instructions and documents about how the survey session was to take place (e.g. how to guarantee confidentiality through predefined ID-codes) as well as a detailed coding manual (e.g. how to handle missing values), including a predefined Excel sheet to enter the data for the written surveys. As a control mechanism, the survey conductors were given the opportunity to go over any issues with the data file when handing in their paper versions. A double check was performed when merging the data into one file. If problems arose during the data collection, they were able to withdraw from this part of the task.

The present study was in line with the general ethical basic assumptions specified in the faculty's general ethical protocol for scientific research. As the protocol states that the advice of the faculty's ethical committee should only be requested in cases of doubt about a research project's conformity to these guidelines, no further ethical approval was necessary.

Instrument

A self-reporting survey instrument was designed with three blocks: (a) demographic information, (b) descriptive questions about PA activity experienced, (c) specific questions about conceptions about PA and its social nature, and anonymity and accuracy within PA. Two different versions of the instrument were created: one with 49 questions concerning PA for students with PA experience, and one with 36 questions concerning PA for students without PA experience. The items in block c of this questionnaire used a 6-point, positively packed rating scale in order to elicit more variance in responses (strongly disagree, mostly disagree, slightly agree, moderately agree, mostly agree, and extremely agree) (Brown, 2004). The following definition of PA was presented to ensure students have a shared understanding: "In a peer assessment-activity students judge each other's tasks/presentations/group assignments. The judgement can be expressed in scores, oral or written feedback or a combination of both." Next, we explain the content of blocks b and c in more detail (see Appendix 1):

a) PA activity description. The first eight questions established an overview of the kind of PA activity students had experienced (block b).
b) Interpersonal processes in PA. The next six questions explored students' beliefs with respect to interpersonal processes within PA (block c).
c) Importance of anonymity in PA. Three items explored the degree to which students considered anonymity to be important in PA, focusing on how much they valued being anonymous (with respect to fellow peers as well as the teacher). (block c)
d) PA accuracy. Two questions were used to establish whether students thought PA was an accurate assessment method (block c).
e) PA conceptions. Three questions explored students' attitudes to PA in terms of usefulness, involvement in the assessment process, and perceived learning gain (block c).

A pilot of this survey was conducted before administration. An expert in formative assessment filled out the questionnaire and that input was used to revise some of the items. The revised survey was then evaluated by a teacher in year 9 and 10 (e.g. comprehension problems, length, etc.). After his input, the survey was then evaluated with a think-aloud procedure by a pupil in grade 10 and one in grade 13. Finally, the survey was piloted with 16 grade 10 students based on all the input from the preceding try-outs. On average, they completed the questionnaire within 20 minutes.

Analysis

To answer RQ1, we calculated the descriptive statistics of the data regarding students' previous PA experience and the PA format. The goal of RQ2 was to understand how students' perception of the educational value of PA was influenced by self-reported beliefs on interpersonal processes, anonymity, and accuracy within PA. Therefore, we identified the relationship of several latent factors and manifest item variables to each other and their contribution to the perceived educational value of PA. For this, structural equation modeling (SEM) was used, as this allowed us to study the fit of a formally defined theoretical model to an empirical dataset. By generating multiple parameters, SEM enabled us to examine each particular hypothesis and incorporate the latent factors that were helpful in removing errors of measurement from the path analysis.

The following fit indices were calculated for every model: first, for *RMSEA*, values between 0 and .06 indicated a very good fit, and values between .06 and .08 indicated a reasonable fit. Second, for *SRMR*, values between 0 and .08 indicated a very good fit. Third, for the *CFI* and *TLI* indices, acceptable values had to be larger than .90, and excellent values had to be above .95. Finally, the χ^2/df (chi squared/degrees of freedom) ratio was considered; for a value to be

considered a good fit, it could not exceed 2.0 (Schreiber, Nora, Stage, Barlow & King, 2006; Schweizer, 2010). We computed the latent factors' reliability through the omega coefficient (McDonald, 1999).

Tests of measurement invariance (configural, metric and scalar) were performed to answer RQ3 about the differences between educational levels, PA experience, and gender. To determine the measurement invariance across subgroups in large samples, it was preferable to report the change in *CFI* and *RMSEA* between the unrestricted and restricted models instead of the difference in chi-squared statistics (Chen, 2007; Cheung & Rensvold, 2002; Meade, Johnson & Brady, 2008; Kline, 2015). As Cheung and Rensvold (2002) recommend using a Δ*CFI* value higher than .01 to indicate a significant drop in fit, and Chen (2007) suggests using Δ*RMSEA* to test for evidence of invariance, the criteria for invariance was Δ*CFI* ≤ .01, Δ*RMSEA* ≤ .015. The tests for latent mean differences were conducted for the groups in which scalar invariance was observed. Assessment of the latent mean differences was based on the critical ratio (*CR*) index, where *CR* ≤ or ≥ 1.96 indicates significant differences in the means. The Cohen's *d* effect size index was also calculated to interpret the magnitude of the mean differences (.20=small differences, .50=medium differences, .80=large differences; Cohen 1988). MPlus 7 (Muthén & Muthén, 2007) was used for the SEM, the tests for measurement invariance, and the tests for latent mean differences. The models' parameters were estimated through robust maximum likelihood (MLR).

Results

RQ1–Do Flemish secondary education students report having experienced PA, and in what format?

Overall, the majority of students in all levels have experienced a PA activity at least once (Table 2). Furthermore, it follows logically that the majority of the students with no experience are situated in Level 1. The frequency of experience of PA was generally high, with "more than thrice" being the most chosen option in Levels 1 (27.76%), 2 (35.15%), and 3 (59.91%).

Table 2
Students' experience with PA

Response category	Level 1 *(*N*=1239)		Level 2 (*N*=1229)		Level 3 (*N*=1105)	
	N	%	*N*	%	*N*	%
None	*263*	21.23	*169*	13.75	*75*	6.79
Once	*209*	16.87	*217*	17.66	*134*	12.13
Twice	*227*	18.32	*223*	18.14	*135*	12.22
Thrice	*196*	15.82	*188*	15.30	*99*	8.96
> Thrice	*344*	27.76	*432*	35.15	*662*	59.91

*Levels: Level 1 = Grade 7-8; Level 2 = Grade 9 -10; Level 3 = Grade 11-12

Looking at the subsample of students with PA experience (*N* = 3066), we can see that PA is most frequently applied in the Dutch, French and math courses (Table 3).

Table 3
Frequency of PA applications in different school subjects

Subject	Level 1* (N=987)		Level 2 (N=1053)		Level 3 (N=1026)	
	N	%	*N*	%	*N*	%
Dutch	*242*	24.5	*144*	13.7	*185*	18.0
English	*61*	6.2	*111*	10.5	*98*	9.6
French	*195*	19.8	*102*	9.7	*150*	14.6
Latin	*24*	2.4	*11*	1.0	*55*	5.4
Greek	*15*	1.5	*N/A*	N/A	*N/A*	N/A
General Sciences	*30*	3.0	*6*	.6	*N/A*	N/A
Math	*157*	15.9	*140*	13.3	*31*	3.0
Economics	*8*	.8	*66*	6.3	*41*	4.0
History	*5*	.5	*60*	5.7	*46*	4.5
Geography	*61*	6.2	*62*	5.9	*27*	2.6
Computer Science	*18*	1.8	*42*	4.0	*31*	3.0
General Projects	*13*	1.3	*76*	7.2	*19*	1.9
Religion	*45*	4.6	*54*	5.1	*61*	5.9

Subject	Level 1* (N=987)		Level 2 (N=1053)		Level 3 (N=1026)	
	N	%	*N*	%	*N*	%
Society Education	*10*	1.0	*55*	5.2	*49*	4.8
Esthetics	*24*	2.4	*10*	.9	*39*	3.8
Professional Courses	*40*	4.1	*33*	3.1	*71*	6.9
Music Education	*28*	2,8	*11*	1.0	*N/A*	N/A
Social Sciences	*11*	1.1	*16*	1.5	*25*	2.4
Physics	*N/A*	N/A	*14*	1.3	*N/A*	N/A
Biology	*N/A*	N/A	*N/A*	N/A	*11*	1.1
Chemistry	*N/A*	N/A	*11*	1.0	*9*	.9
Gym	*N/A*	N/A	*29*	2.8	*7*	.7
Spanish	*N/A*	N/A	*N/A*	N/A	*10*	1.0
German	*N/A*	N/A	*N/A*	N/A	*61*	5.9

* Levels: Level 1 = Grade 7-8; Level 2 = Grade 9 -10; Level 3 = Grade 11-12

The majority of students in all levels mentioned that the PA activities they experienced were not typically, or only sometimes, taken into account for the monthly report in terms of grades (see Table 4). Together with the fact that the PA activities took place during a lesson, it is possible that the current use of PA in Flemish secondary education is formative in nature (although this is a highly speculative explanation).

Regarding student training to perform PA, the current findings are in high contrast with the current research guidelines: on all educational levels (74.61% for Level 1, 78.76% for Level 2, and 80.66% for Level 3) students reported not having received PA training. On the other hand, two-thirds of the student population with PA experience reported that they were involved in defining the PA criteria. Surprisingly, the percentage of active involvement decreases from Level 1 (50.64%) to Level 3 (38.83%).

Table 4
Format of PA-usage for students with PA experience

Question and response category	Level 1 (*N*=957)		Level 2 (*N*=1015)		Level 3 (*N*=1006)	
	N	%	*N*	%	*N*	%
Grades: Was the result of the PA activity mentioned on the monthly report?						
Yes	*140*	14.63	*210*	20.69	*273*	27.14
No	*444*	46.39	*402*	39.61	*286*	28.43
Sometimes	*373*	38.98	*403*	39.70	*447*	44.43
Time: The PA activity took place						
…during a lesson	*574*	59.54	*552*	54.33	*470*	48.25
…at the end of series of lessons	*390 (N= 964)*	40.46	*464*	45.67	*504*	51.75
Training: Were you trained to perform PA?						
Yes	*128*	13.43	*86*	8.30	*75*	7.48
No	*711*	74.61	*816*	78.76	*809*	80.66
Sometimes	*114*	11.96	*134*	12.93	*119*	11.86
Involvement in defining PA criteria: Were you involved in defining the PA criteria?						
Yes	*593*	50.64	*453*	43.77	*393*	38.83
No	*345*	29.46	*231*	22.32	*302*	29.84
Sometimes	*233*	19.90	*351*	33.91	*317*	31.32

Table 5 provides details of the students' responses on guaranteed anonymity within the PA activities experienced. Generally, anonymity was not provided for the assessor, the assessees, nor the teacher. More specifically, if anonymity was provided, anonymity for the assessor was the most used anonymous mode. This usage of this anonymous mode was reported to be 10 to 15% at all educational levels. The fact that about one-third indicated "sometimes" as the chosen option in all anonymous modes possibly indicates that teachers are exploring with different anonymous modes depending on the nature of the PA activity.

Table 5
Anonymity within PA activity

Response category	Level 1 (*N*=976)		Level 2 (*N*=1040)		Level 3 (*N*=1010)	
	N	%	*N*	%	*N*	%
Anonymity for the assessor						
Yes	*111*	*11.48*	*170*	*14.57*	*148*	*14.65*
No	*576*	*59.57*	*573*	*49.10*	*548*	*54.26*
Sometimes	*289*	*29.89*	*297*	*25.45*	*314*	*31.09*
Anonymity for the assessee						
Yes	*57*	*5.83*	*44*	*4.25*	*33*	*3.26*
No	*801*	*81.99*	*887*	*85.70*	*879*	*86.94*
Sometimes	*119*	*12.18*	*104*	*10.05*	*99*	*9.79*
Anonymity for the teacher						
Yes	*80*	*8.34*	*56*	*5.38*	*45*	*4.45*
No	*658*	*68.61*	*821*	*78.87*	*839*	*82.99*
Sometimes	*221*	*23.04*	*164*	*15.75*	*127*	*12.56*

RQ2 – What variables predict students' PA conceptions?

To study the latent constructs and their relationships, we fitted a structural equation model. In this model, we proposed five latent constructs or factors to explain the responses to the survey's items. This allowed us to study the relationships between the constructs at the latent level. For this, we used only the questions from block c in our instrument. The first factor is the "*educational value of PA*" and refers to the extent to which students think PA is a valuable assessment method. The second factor is "*trust in evaluative capabilities*," which measures the extent to which students believe in their own and their peers' evaluative capabilities. The third factor is "*negative interpersonal processes*," which refers to the extent to which students deem that friendship marking is going on and the amount they fear disapproval if they give low scores and/or critical feedback. The fourth factor is the "*importance of anonymity*," which evaluates the amount of importance attributed to guaranteeing anonymity to assessors, assessees and/or teachers within PA activities. The fifth factor is "*accuracy*," which measures students' perceived accuracy of PA and whether they think their peers are capable of giving accurate judgments. In order to correctly identify the model, the variance of all five latent variables was fixed to one. Since we wanted to know what variables predict the "*educational value of PA*," we forced this latent construct to be

explained (i.e. receive loadings) by the other four. Therefore, the model is conceptually equivalent to a multiple regression with one latent dependent variable and four latent predictors. Five items show low factor loadings, which indicates a low communality with the rest of the scale. We decided to remove these items in order to gain conceptual clarity. The full SEM model (with standardized loadings) is shown in Figure 1.

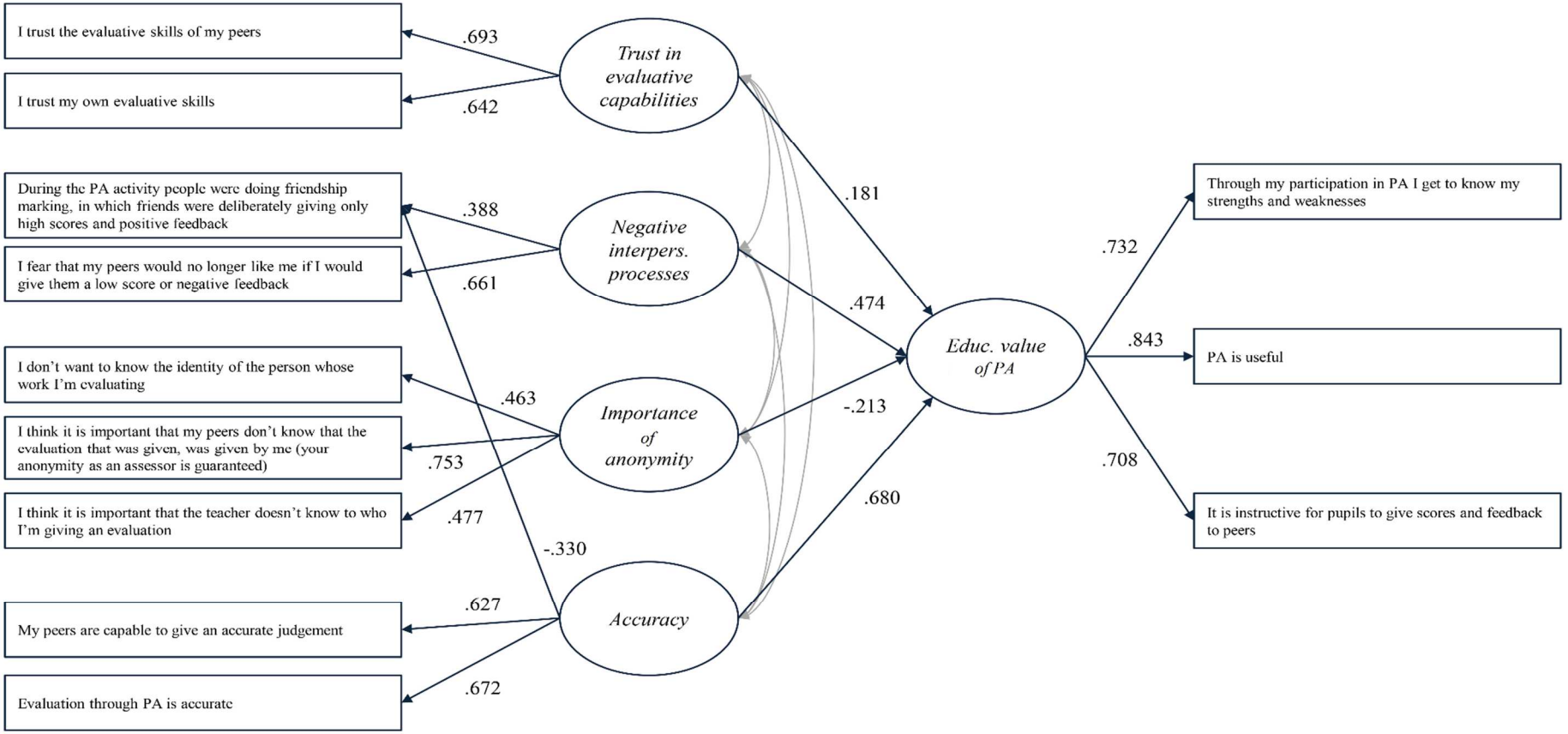


Figure 1. Baseline SEM model with standardized loadings*

* *Correlations among the latent predictors reported in Table 6 for clarity*

Since our instrument taps several conceptually different constructs, each of them contributing to an overall *formative index* (*Educational value towards PA*), it is not recommended to evaluate the instrument's reliability through one internal consistency coefficient such as Cronbach's alpha (Streiner, 2003; Sijtsma, 2009). Instead, we assessed the reliability of our measures by computing the omega coefficient for each of the latent factors. This strategy is often recommended for assessing reliability when a latent factor structure is available (McDonald, 1999). The omega value for *Trust in evaluative capabilities* was ω = 0.62, for *Negative interpersonal processes* ω = 0.41, for *Importance towards anonymity* ω = 0.60, for *Accuracy* ω = 0.50, and for *Educational value towards PA* ω = 0.44. These values, although not very high, are acceptable considering that each factor was measured by a very reduced number of observable indicators.

The correlations among the four latent predictors are shown in Table 6. The total amount of variance for the "*educational value of PA*" explained by the predictors is R^2 = .548 (*std. error* = .046). Note that the latent predictors for "*trust in evaluative capabilities*" (β = .181, *std. error* = .081, *p* = .022), "*negative interpersonal processes*" (β = .474, *std. error* = .126, *p* < .001), and "*accuracy*" (β = .680, *std. error* = .038, *p* < .001) had positive loadings on the "*educational value of PA*," meaning that higher values in the former tend to be associated with higher values in the latter. Hypothesis H2a was therefore rejected and H2b partly accepted as only trust was identified as a positive interpersonal variable in the model. The identification of "*accuracy*" as a positive predictor confirms hypothesis H2c. On the other hand, in line with hypothesis H2d, "*Importance of anonymity*" (β = -.213, *std. error* = .077, *p* = .005) appears to be a negative predictor of students' "*educational value of PA.*" All four regression loadings are statistically significant.

This model achieves an acceptable to good fit: χ^2 / df = 120 / 43 = 2.79 (note that the chi squared values are highly inflated by large sample sizes such as ours), *CFI* = .987, *TLI* = .980, *SRMR* = .018 and *RMSEA*= .024, with a 90% confidence interval between .019 and .029. All the items load significantly onto their corresponding latent factors (*p* < .001). The item loadings range between -.33 and .84.

Table 6
Correlation among latent predictors

$N = 3044$ $p < 0.001$ in all cases	Neg. interp. processes	Importance of anonymity	Accuracy
Trust in evaluative capabilities	-.700	-.498	.490
Negative interpers. processes		.692	-.318
Importance of anonymity			-.298

RQ3 – What differences are there in the relations among students' perceptions of interpersonal variables, perceived accuracy, the attributed importance of anonymity and the perceived educational value of PA that exist between gender, educational levels, and students with different PA experiences?

In order to enable a comparison of the mean scores of the latent constructs across educational levels, differing PA experiences and gender, first we tested the scalar invariance between the different groups. If scalar invariance holds, differences in the means of the observed items can be interpreted as a consequence of the differences in the means of the latent constructs.

Table 7
Results of the tests of measurement invariance by (1) Level, (2) PA Experience, and (3) Gender

Multi Group CFA		Model Fit Indices				Model Comparisons		
		X^2 (df)	CFI	TLI	RMSEA	Comparison	ΔCFI	ΔRMSEA
(1) Level	Model 1 (configural invariance)	224.0 (129)*	.984	.976	.027			
	Model 2 (metric invariance)	267.3 (155)*	.980	.974	.028	Model 2 vs. Model 1	-.004	.001
	Model 3 (scalar invariance)	327.8 (169)*	.974	.969	.030	Model 3 vs. Model 2	-.006	.002
(2) PA Experience	Model 1 (configural invariance)	246.2 (173)*	.987	.981	.024			
	Model 2 (metric invariance)	302.5 (211)*	.984	.980	.024	Model 2 vs. Model 1	-.003	.000
	Model 3 (scalar invariance)	322.7 (232)*	.984	.982	.023	Model 3 vs. Model 2	.000	-.001
(3) Gender	Model 1 (configural invariance)	179.4 (86)*	.984	.976	.027			
	Model 2 (metric invariance)	194.6 (99)*	.984	.978	.025	Model 2 vs. Model 1	.000	-.002
	Model 3 (scalar invariance)	247.4 (106)*	.976	.970	.030	Model 3 vs. Model 2	-.008	-.005

* <.05 / With (1) 987 students in Level 1, 1,053 students in Level 2 and 1,026 students in Level 3; (2) 561 students with one PA experience, 586 with two, 481 with three and 1,438 with more than three; (3) 1,675 girls and 1,389 boys – the criteria for invariance model comparison: ΔCFI ≤.01, ΔRMSEA ≤.015

As Table 7 shows, scalar invariance holds for all three comparisons, i.e. comparing educational level, PA experience and gender. This means that students across the different groups interpreted our measurement instrument in a consistent way and that the mean scores of the latent constructs can be compared.

After scalar invariance was verified, we studied the differences between latent means. To do so, the mean of one of the groups (the reference group) is fixed to zero in all latent variables, so the means of the other groups can be freely estimated and interpreted as a difference between those groups and the reference group. In the present study, students from Level 1 (1), with only one PA experience (2), and males (3) were the reference groups in each comparison. Note that, since the metric of all latent variables is arbitrary, the comparisons of latent mean differences do not enable the estimation of the absolute mean in each group but rather study the differences in the latent variables among groups. Table 8 shows the results of the test for the latent mean differences for each latent variable. Since we fixed all the variances to one, the estimated mean differences can be interpreted as Cohen's standardized differences (Cohen, 1988).

Table 8
Test of latent mean differences

	Latent construct	*d*	*S.E. for difference*	*CR* (*z* = Estimate / S.E.)	*p*	*d*	*S.E. for difference*	*CR* (*z* = Estimate / S.E.)	*p*	*d*	*S.E. for difference*	*CR* (*z* = Estimate / S.E.)	*p*
			Second vs. First Level				Third vs. First Level						
(Comparison 1) Level	Trust in evaluative capabilities	.025	.057	.432	.666	-.182	.058	-3.124	.002				
	Negative interpersonal processes	.007	.068	.107	.915	-.100	.066	-1.523	.128				
	Importance of anonymity	.027	.057	.471	.638	-.016	.057	-.289	.773				
	Accuracy	-.335	.060	-5.619	<.001	-.588	.061	-9.654	<.001				
	PA Educ. Value	.014	.047	.295	.384	-.251	.046	-5.504	<.001				
			Twice vs. Once				Thrice vs. Once				More than thrice vs. once		
(Comparison 2) PA Experience	Trust in evaluative capabilities	.144	.078	1.858	.063	.150	.084	1.776	.076	.281	.066	4.235	<.001
	Negative interpersonal processes	.026	.082	.321	.748	.104	.091	1.141	.254	-.025	.071	-.357	.721
	Importance of anonymity	-.131	.074	-1.776	.076	-.182	.080	-2.286	.022	-.235	.064	-3.648	<.001
	Accuracy	.070	.075	.941	.347	.150	.079	1.898	.058	.122	.063	1.952	.051
	PA Educ. Value	.280	.061	4.585	<.001	.318	.068	4.659	<.001	.488	.040	12.244	<.001
			Females vs. Males										
(Comparison 3) Gender	Trust in evaluative capabilities	.024	.046	.506	.613								
	Negative interpersonal processes	.336	.053	6.285	<.001								
	Importance of anonymity	.110	.046	2.402	.016								
	Accuracy	.047	.047	1.009	.313								
	PA Educ. Value	.177	.039	4.495	<.001								

Students in Level 2 (n = 1046) and Level 3 (n = 1017) have a significantly lower score for "*accuracy*" compared to students in Level 1 (n = 981). The effect sizes of the mean differences are, however, moderate, being d = .34 and d = .59, respectively. Students in Level 3 show a significantly lower mean in "*trust in evaluative capabilities*" than students in Level 1. Nevertheless, the effect size is small (d = .18).

Students that have experienced PA three times (n = 478) attribute significantly less "*importance of anonymity within PA*" than students with only one experience (n = 554) (small effect d = .18). Students with more than three PA experiences (n = 1430) have a significantly lower mean concerning "*importance of anonymity within PA*" than the once group (small effect d = .28), and a significantly higher mean in "*trust in evaluative capabilities*" than the once group (small effect d = .24).

Girls (n = 1665) show a significantly higher mean in "*negative interpersonal processes*" (d = .34) and the "*importance of anonymity within PA*" compared to boys.

Regarding the attributed "*educational value of PA*", students in the Level 3 attribute a significantly lower value to PA than students in the Level 1 (d = .25). On the contrary, the more experience students have, the more positive attitude toward PA they report (d = .28; d = .32; d = .49). Overall, females report a higher educational value of PA compared to males (d = .18). All the remaining comparisons were not significant.

Discussion

This cross-sectional survey study explored Flemish secondary-education students' self-reported experiences with PA, with a specific focus on the inherent social nature of the activity. The majority of this 3,066-student sample had previous experience with PA activities. This is in high contrast with previous studies (Noonan and Duncan, 2005; Panadero et al. 2016). More specifically, here, PA was frequently used in all levels of secondary education, with the majority of students indicating that they have experienced it multiple times. The nature of the PA activities was relatively similar: most PA activities took place during a lesson and, if grades were awarded, they were not accounted for in a summative manner. This possibly mirrors a formative PA approach.

Students were, in most cases, not trained to perform PA, which goes against the frequently mentioned PA practice guidelines (Sluijsmans, 2002; Panadero et al., 2016). However, students on all levels were frequently involved in defining assessment criteria, which is an assessment scaffold that previous research has shown to be conducive for the overall quality of PA activities, as it clarifies expectations and recognizes students as valuable and active actors in the assessment

process (Panadero & Romero, 2013). Regarding anonymity, the great majority of students indicated that PA activities were conducted in a non-anonymous mode.

In this work, we used SEM as a tool for understanding the relationship between the latent constructs of interpersonal variables, anonymity, perceived accuracy and the value students attribute to PA. This approach allowed us to find relevant relations between the studied constructs, and to study the differences between the groups. However, it should be noted that this is the first large-sample-based exploration of this under-studied problem, and the model presented here should not be considered to be a universal model. More research is needed with samples from different countries and ages to check whether the latent structure and differences found here are consistently found in other scenarios. Our results lead us to believe that there are four latent factors that are relevant to predicting the educational value students attribute to PA: (1) *trust in evaluative capabilities*, (2) *negative interpersonal processes*, (3) *accuracy* and (4) *importance of anonymity within PA.* In contrast to a previous study by van Gennip et al. (2010), "*value congruency*" and "*psychological safety*" were not needed to explain the students' conceptions. A possible reason for this could be the fact that, in Van Gennips' (2010) quasi-experimental design, students' perceptions were measured directly after the intervention they conducted, causing a stronger effect from these variables. Another possible explanation could be that our instrument failed to tap the aspects of *Value Congruency* and *Psychological Safety* that are relevant for the perceived *Educational Value*. According to our multiple regression SEM model, 54.8% of the variance in the "*educational value of PA*" can be explained solely by the other four factors. We conducted several tests for measurement invariance, providing support for scalar invariance for all groups of interest; that is, educational level, PA experience, and gender. The analysis revealed that the three factors (1) *trust in evaluative capabilities*, (2) *negative interpersonal processes*, and (3) *accuracy* were positive contributors to students' perceived educational value of PA.

In line with previous research, reciprocal trust in the assessment skills and capability of peers to give judgement on your work has proven to be important within the PA process (van Gennip et al., 2010), as it may lead to deeper learning approaches (Cheng & Tsai, 2012). Our model showed that the amount of "*trust in your own and peers' evaluative capabilities*" positively contributes to students' perceived educational value of PA. However, the fact that students in Level 3 had a significantly lower amount of trust compared to Level 1 raises concerns. A possible explanation for this is that the current PA implementation in the higher levels of education is still predominantly aimed at scoring, although it is not summatively accounted for, instead of pursuing a "deep" approach with much more attention to the peer feedback component, which was also reported by Panadero & Brown (2016). This might also explain the significantly lower amount of "*educational value of PA*" between Levels 3 and 1.

"*Negative interpersonal processes*" were found to be a positive predictor of students' perceived "*educational value of PA*." Possibly the items in our questionnaire were posed rather descriptively, in the sense that they rather indirectly refer to a value concept stronger than "fear" and "unfairness" as currently used. However, as the correlations between the latent predictors indicate a correct interpretation of the items, it is up to future research to deepen our understanding of the interrelationship between these variables. As a consequence, the current findings suggest that students' awareness levels about the fact that these processes–including their potential undesirable effects–are possibly present in PA, leads to placing greater value on peer assessment as valuable learning activity. The test of mean latent differences revealed that girls' ratings are significantly higher on this variable.

Perceived "*accuracy*" proved to be the most important positive predictor in the structural model. However, as students in Levels 2 and 3 reported significantly lower rates on perceived accuracy compared to Level 1, again a predominant approach to summative scoring alone instead of stimulating interactive feedback processes could be an explanation for this finding. One possible way to counter this decrease might be to consequently train students how to participate in PA tasks in order to increase the (perceived) accuracy even when they are in the higher levels of secondary education.

Regarding the relation to students' "*educational value of PA*" and the factor "*importance of anonymity*," a negative relationship was found. This confirms the previous theoretical work on PA by Topping (1998) who indicates that privacy is an important structural component of PA, and confirms the findings of small-scale studies by, for example, Raes et al. (2013) and Vanderhoven et al. (2015) who found that students prefer the anonymous modes of PA within face-to-face settings. The use of anonymous modes of PA was in general very low in the studied sample. Apparently, Flemish teachers are either not aware that the importance attributed to anonymity can influence the outcomes of a PA activity (Vanderhoven et al., 2015) or, as found by Panadero et al. (2016), teachers believe in using anonymous PA formats, but currently do not implement such formats. The mean differences between the experience categories revealed that students' "*importance of anonymity*" was lower for students with three or more than three PA experiences. This finding suggests that practice leads to more willingness to participate in non-anonymous PA settings. In these settings, assessors and assessees can interact and, preferably, provide richer feedback; i.e. more negative verifications, more information about how to correct mistakes, and suggestions for how to improve their work. Interestingly, girls attributed higher importance to anonymity. As a consequence, one could advocate that offering anonymity can be used as a temporary catalyst to a) accelerate the creation of a safe PA environment, and b) acknowledge the time that students need to get accustomed to the interpersonal processes that are evoked through

participation in PA and that might possibly be mitigated through anonymity (Vanderhoven et al., 2015).

Implications

This study adds to our understanding of the impact of students' perceptions on interpersonal processes, accuracy, and anonymity, and their relation with the perceived educational value of PA. It has important consequences for educational practices. First of all, trust building in their own and others' evaluative capabilities should be seen as an essential step when implementing PA activities. Second, teachers should discuss the inherent social nature of PA (Panadero, 2016), its educational value, and the potential negative interpersonal processes (e.g. friendship marking, hostile behavior) before the start of the actual PA task in order to raise students' awareness of these issues because they positively influence students' PA conceptions. Moreover, teachers should give suggestions for how to counter these negative effects; for example, by exploring the importance attributed to anonymity by students, and create a safe learning environment in their classrooms. Third, it is important that teachers with mostly girls in their classes pay extra attention to the creation of a positive atmosphere that counters negative processes. Additionally, as girls seem to prefer anonymous PA modes, active guidance for more interactive, non-anonymous modes might be needed.

It is up to future research to deepen our insights on these topics, but it is clear that the implementation of different assessment scaffolds (e.g. active involvement in criteria development, training, and trust building) to improve the educational potential of PA will take a considerable amount of time. Therefore, thoughtful implementation of PA by teachers is needed. Evidently, as also advocated by Brown and Panadero (2016), pre-service and in-service teachers should be trained in PA implementation and, more importantly, have repeatedly experienced this for themselves, in order to achieve insight on the effects of the interpersonal processes that are inherent to this assessment method.

Limitations and Future Lines of Research

This cross-sectional study's major limitation is its self-reported nature. This implies that students' answers may have been influenced by social desirability, as this is a risk with any form of subjective data collection (Desimone, 2009). However, throughout the process of survey development and administration, several steps were taken to reduce social desirability bias. This included extensive piloting, critical reviews, and pretesting by an expert in formative assessment. Moreover, confidentiality for respondents was assured. Another matter of concern is the vague quantity "sometimes" in the response scale of the items regarding RQ1. Evidently, there are

possible memory problems within individuals when recalling what specific characteristics the PA activities they have experienced, and there is variability between individuals in how such a vague frequency is understood.

It is reasonable to think that some important aspects of the latent constructs were not adequately tapped by the questions in our instrument. Our study involved complex psychological constructs, and the model included a reduced number of items for measuring each one. This allowed us to measure the latent factors with only moderate reliability, as shown by the omega values. This may also be the reason why *Psychological Safety* and *Value Congruency* did not appear to be relevant and were discarded in our SEM model. Nevertheless, we succeeded to explain 54.8% of the variance in *Educational Value of PA* basing on the other four factors in our final model. This leads us to think that the four "predictive" latent factors are relevant to explain students' educational value towards PA. In other words, our model was able to explain the differences in this relevant construct to a large extent, even with such a reduced number of items. This is consistent with previous research showing that complex psychological constructs can be measured through very short scales (Rammstedt & John, 2007). In any case, it would be interesting to include broader measures of these same constructs in future studies.

In this study we did not take into account students' general conceptions of assessment, and their relation or impact on our structural model, although a reciprocal and influencing relationship between them is plausible. It is up to future studies to look into this so far under-researched relationship. Additionally, this survey study did not take into account the teachers' opinions on the researched variables. Again, future research could focus on the linkage between the assessment and PA perceptions of these two actors.

Conclusions

Even though the interest in the inherent social nature of PA processes has increased over the past decade (e.g. Panadero, 2016), most of the experimental studies have been small-scale studies. This large-sample survey study has confirmed the complex nature of PA processes, which can trigger powerful feelings in our students and have an impact on their beliefs in the educational value of PA. The four identified predictors of students' perceived educational value of PA can be of guidance to program developers, instructional designers, teacher educators, and teachers when designing and implementing PA tasks. The results confirm Panadero's (2016) theoretical work on the social nature of PA: PA does not happen in a vacuum and a shallow implementation of it might do more harm than good. Mitigating the interpersonal variables will ask for intensive, repeated, highly interactive PA tasks in which interpersonal processes are actively monitored. Additionally, sufficient classroom time for the improvement of students' peer feedback skills should be made.

Therefore, the structural integration of PA activities within the curriculum is strongly recommended.

References

Ajzen, I. (2005). *Attitudes, Personality, and Behavior* (2nd ed.). New York: Open University Press.

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing Procedures for Implementing Peer Assessment in Large Classes Using an Action Research Process. *Assessment & Evaluation in Higher Education*, *27*(5), 427–441. doi:10.1080/0260293022000009302

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. doi:10.1080/0969595980050102

Bolzer, M., Strijbos, J. W., & Fischer, F. (2015). Inferring mindful cognitive-processing of peer-feedback via eye-tracking: Role of feedback-characteristics, fixation-durations and transitions. *Journal of Computer Assisted Learning*, *31*(5), 422–434. doi:10.1111/jcal.12091

Boud, D. (1995). Assessment and learning: Contradictory or complementary? In K. Peter (Ed.), (pp. 35–48). London: Kogan Page.

Brown, G., & Harris, L. (2011). Student conceptions of assessment by level of schooling: Further Evidence for ecological rationality in belief systems. *Australian Journal of Educational & Developmental Psychology*, *12*, 46–59.

Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice*, *15*(1), 3–17. doi:10.1080/09695940701876003

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, *27*(1), 210–220. doi:10.1016/j.tate.2010.08.003

Brown, G. T. L., McInerney, D. M., & Liem, A. D. (2009). Students perspectives of assessment: Considering what assessment means to learners. In D. M. McInerney, G. T. L. Brown, & A. D. Liem (Eds.), *Student Perspectives on Assessment: What Students Can Tell Us about Assessment for Learning* (pp. 1–21). Charlotte, NC: Information Age Publishing.

Carless, D. (2013). Trust and its role in facilitating dialogic feedback. In *Feedback in Higher and Professional Education* (pp. 90–103). doi:10.4324/9780203074336

Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education*, *35*(5), 551–564. doi:10.1080/02602931003632381

Cheng, K. H., & Tsai, C. C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology*, *28*, 599–618.

Cowie, B. (2009). My teacher and my friends helped me learn: Student perceptions and experiences of classroom assessment. In D. M. McInerney, G. T. L. Brown, & A. D. Liem (Eds.), *Student Perspectives on Assessment: What Students Can Tell Us about Assessment for Learning* (pp. pp. 85 – 105). Charlotte, NC: Information Age Publishing.

Cowie, B., Moreland, J., & Otrel-Cass, K. (2013). *Expanding Notions of Assessment for Learning: Inside Science and Technology Primary Classrooms*. Springer Science & Business Media.

De Backer, F., & Philips, I. (2013). *CTO/SDL Toolkit Competenties Nederlands Breed Evalueren*. Retrieved from http://www.ond.vlaanderen.be/toetsenvoorscholen/toolkit_breed_evalueren/pdf/ToolkitInZijnGeheel.pdf

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

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, *24*, 331–350. doi:10.1080/03075079912331379935

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287–322. doi:10.3102/00346543070003287

Flemish Goverment. (2015). *Flemish Education in Figures*. Retrieved from https://www.vlaanderen.be/nl/publicaties/detail/flemish-education-in-figures-2014-2015

Ghent University – Department of Educatoinal Studies. (2013). *Wiskundige geletterdheid bij 15-jarigen - Overzicht van de eerste Vlaamse resultaten van PISA2012*. Retrieved from http://www.pisa.ugent.be/uploads/assets/106/1396273183438-KORTE BROCHURE PISA2012.pdf

GO! (2012). *Evaluaren in het secundair onderwijs*. Retrieved from http://pro.g-o.be/blog/Documents/GO!_Visietekst_evalueren in het SO_DEF.pdf

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, *36*, 101–111. doi:10.1016/j.tate.2013.07.008

Harris, L. R., Brown, G. T. L., & Harnett, J. A. (2014). Understanding classroom feedback practices: A study of New Zealand student experiences, perceptions, and emotional responses. *Educational Assessment, Evaluation and Accountability*, *26*(2), 107–133. doi:10.1007/s11092-013-9187-5

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, *38*, 21–27. doi:10.1016/j.stueduc.2012.04.001

Howard, C. D., Barrett, A. F., & Frick, T. W. (2010). Anonymity to promote peer feedback: Pre-service teachers' comments in asynchronous computer-mediated communication. *Journal of Educational Computing Research*, *43*(1), 89–112. doi:10.2190/EC.43.1.f

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling*. Guilford Publications.

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, *35*(5), 501–517. doi:10.1080/02602931003786559

Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research & Evaluation*, *10*(17).

McDonald, R. P. (1999). *Test theory: a unified treatment*. Mahwah, N.J: L. Erlbaum Associates.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*, 568. doi: 10.1037/0021-9010.93.3.568

Muthén, L., & Muthén, B. (2007). *Mplus user's guide (version 7)*.

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Human factors and social conditions of assessment.* (pp. 1–39). New York, NY: Routledge.

Panadero, E., & Brown, G. T. L. (2016). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education*., 32:133 doi:10.1007/s10212-015-0282-5

Panadero, E., Jonsson, A., & Strijbos, J. W. (2016). Enhancing ownership, relinquishing control: Scaffolding students regulated learning through involvement via assessment for learning. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the Challenge of Implementation.*

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 133–148. doi:10.1080/0969594X.2013.877872

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, *39*(4), 195–203. doi:10.1016/j.stueduc.2013.10.005

Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality,

41, 203-212

Reinholz, D. (2015). The assessment cycle: A model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 1–15. doi:10.1080/02602938.2015.1008982

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*, 323-338. doi:10.3200/JOER.99.6.323-338

Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment, 26*, 1-2. doi:10.1027/1015-5759/a000001

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika, 74*, 107. doi:10.1007/s11336-008-9101-0

Sluijsmans, D. M. A. (2002). Student involvment in assessment: The training of peer assessment skills. (Unpublished doctoral dissertation). Open University of the Netherlands, Heerlen.

Segers, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation*, *37*(1), 49–54. doi:10.1016/j.stueduc.2011.03.008

Streiner, D. L. (2003). Being Inconsistent About Consistency: When Coefficient Alpha Does and Doesn't Matter. *Journal of Personality Assessment, 80*, 217-222. doi:10.1207/S15327752JPA8003_01

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, *30*(4), 325–341. doi:10.1080/02602930500099102

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimizing new modes of assessment: In search of qualities and standards.* (pp. 55–87). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*(4), 339–343. doi:10.1016/j.learninstruc.2009.08.003

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, *4*, 41–54. doi:10.1016/j.edurev.2008.11.002

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, *20*, 280–290. doi:10.1016/j.learninstruc.2009.08.010

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, *81*, 123–132. doi:10.1016/j.compedu.2014.10.001

Vickerman, P. (2009). Student perspectives on formative peer assessment: An attempt to deepen learning? *Assessment & Evaluation in Higher Education*, *34*, 221–230. doi:10.1080/02602930801955986

Vlaanderen, K. O. (2013). *Evalueren en leren: de kip of het ei?*, 10–13. Retrieved from https://pincette.vsko.be/VSKO/Documentatiedienst/Tijdschriften/Breedbeeld/Publiek/2012-2013/Breedbeeld 2012-2013-3 Dossier.pdf

Yu, F.Y., & Sung, S. (2015). A mixed methods approach to the assessor's targeting behavior during online peer assessment: Effects of anonymity and underlying reasons. *Interactive Learning Environments*, 1–18. doi:10.1080/10494820.2015.1041405

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*. Retrieved from https://books.google.com/books?hl=nl&lr=&id=og4hVOcjcqMC&pgis=1

Appendix 1

Students' conceptions about PA

Definition: In a PA-activity, students judge each other's tasks/presentations/group assignments. The judgement can be expressed as scores, oral or written feedback, or a combination of both.

#	Question	Response Format
1	Have you ever experienced PA?	Once–Twice–Thrice–> Thrice–None
2	Via what tool the PA was conducted?	Paper–Oral–Computer–Combination
3	Did the results of the PA account for the monthly report?	Y/N/Sometimes
4	At what moment did the PA activity take place?	During / At the end of the series of lessons
5	Anonymity was guaranteed for: a) Assessors b) Assessees c) Teachers	Y/N/Sometimes, in each of three categories
6	Were you trained to do PA?	Y/N/Sometimes
7	Were you involved in defining the assessment criteria?	Y/N/Sometimes
8	Was the result of the PA orally discussed with the teacher?	Y/N/Sometimes
9i	Negative interpersonal–friendship marking: During the PA activity, people were doing friendship marking, in which friends were deliberately giving only high scores and positive feedback	6 point Likert-scale (2 negatively formulated, 4 positively formulated)
9ii	Negative interpersonal–fear of disapproval: I fear that my peers would no longer like me if I give them a low score or critical feedback.	Idem

9iii	Interpersonal trust peer: I trust the evaluative skills of my peers	Idem
9iv	Interpersonal trust own: I trust my own evaluative skills.	Idem
9v	Interpersonal psychological safety: I feel comfortable giving my opinion on a peer's work for the whole class group.	Idem
9vi	Interpersonal value congruency: Everyone was interpreting the evaluation criteria in the right way.	Idem
10i	Anonymity: I think it is important that my peers don't know that the evaluation that was given, was given by me*. * Your anonymity as an assessor is guaranteed.	Idem
10ii	Anonymity: I don't want to know the identity of the person whose work I'm evaluating.	Idem
10iii	Anonymity: I think it is important that the teacher doesn't know to who I'm giving an evaluation.	Idem
11i	Accuracy: My peers are capable of giving an accurate judgement.	Idem
11ii	Accuracy: Evaluation through PA is accurate.	Idem
12i	PA conceptions: Through my participation in PA, I get to know my strengths and weaknesses.	Idem
12ii	PA conceptions: PA is useful.	Idem
12iii	PA conceptions: It is instructive for pupils to give scores and feedback to peers.	Idem

3 How do teachers perceive the educational value of PA in relation to its social nature?

This chapter is based on:

Rotsaert, T., Schellens, T., & Panadero, E. (submitted). How do teachers perceive the educational value of peer assessment in relation to its social nature? Manuscript under review for publication in *Studies in Educational Evaluation*

Chapter 3
How do Teachers Perceive the Educational Value of Peer Assessment in Relation to Its Social Nature?

Abstract

This study explores teachers' use of peer assessment (PA) as well as their level of awareness regarding students' considerations of the interpersonal processes within PA. Furthermore, it examines how these perceptions relate to the educational value teachers attribute to PA. A quantitative survey study (N = 225) was conducted in secondary schools in Flanders, which is a context characterized by a high self-governance policy and low demands for external assessment accountability. Results show that teachers are moderately aware with regard to their students' concerns about the impact of interpersonal processes in PA as well as the importance students attribute to anonymity within PA. However, no significant relationship for the educational value teachers attributed to PA was found for these variables. This study illustrates that teachers' perceived accuracy of PA is a major predictor of their belief in its educational value.

Introduction

Peer Assessment (PA) has been shown to have positive effects on students' motivation and engagement in learning (Topping, 2003). In this collaborative classroom situation, peers use one another as a resource, both by sharing ideas and evaluating the ideas of others, and by providing peer feedback (PF), which can be quantitative (e.g. grades or ratings across assessment criteria) and/or qualitative (e.g. written or oral comments) (Bolzer, Strijbos, & Fischer, 2015; Topping, 2010). PA offers many benefits (Dochy, Segers, & Sluijsmans, 1999; Reinholz, 2015) and can be conceived of as a learning tool due to the active involvement of the learner in the learning process and its provision of skills to the learner in order to assess the criteria that define high-quality work (Reinholz, 2015; Topping, 1998). Furthermore, as an assessment tool, PA is a strong vehicle of 'assessment for learning' because it actively involves students in evaluating their learning and allows them to participate in collaborative appraisal through the use of multiple perspectives when incorporating viewpoints from different learners (Falchikov, 2003; Panadero, 2016).

Despite its benefits, students' interpersonal perceptions can influence their learning from PA (Cowie & Harrison, 2016). For example, students may experience pressure due to friendships with peers (Raes, Vanderhoven, & Schellens, 2013). Until recently, the social nature of this specific

classroom assessment method and its impact on students' learning has only been explored in small-scale intervention studies, mainly within vocational and higher education contexts and with an emphasis on the students' perspectives (Panadero, 2016; van Gennip, Segers, & Tillema, 2010). As a consequence, there is a need not only to explore this phenomenon on a larger scale, but also to obtain an insight into teachers' levels of awareness regarding their students' understandings of the interpersonal process within PA. This is important, as such an insight might be a determining factor for teachers' implementation of PA in their daily classroom practice as well as having the potential to inform future professional development initiatives.

What do We Know about Teachers' PA Conceptions?

The teacher's perspective and actions are critical in developing a culture of classroom assessment that supports the sharing of ideas and explanations in a manner that is accepting of individual differences (Cowie & Harrison, 2016; Harris & Brown, 2013). Therefore, studying teachers' conceptions of peer assessment is important at a time when the innovation of assessment practices is on the educational agenda (Brown & Harris, 2016).

Previous research about teachers' conceptions of PA use in the classroom shows that teachers value PA as a learning activity, but that it is only used occasionally (Noonan & Duncan, 2005). Panadero and Brown (2017), in a recent survey-study of Spanish teachers, came to the conclusion that, although overall teachers like the instructional use of PA, they sometimes struggle with its inherent difficulties (e.g. students' possible lack of objectivity) and only use PA if previous experiences have been positive. Interestingly, primary and secondary school teachers reported higher levels of PA implementation and certainty about its educational value, in contrast with higher education teachers (Panadero & Brown, 2017). Similar results were found in two previous studies investigating how higher education and secondary school teachers perceived PA (Lynch & Golen, 1992; Noonan & Duncan, 2005). Although the aforementioned studies help us to better understand the effect of teachers' beliefs on the self-reported used of this assessment practice, these studies did not focus on the human and social conditions that can stimulate and/or impede the enactment of valuable PA practices.

Due to PA's emotionally charged nature (for a review, see Panadero, 2016), teachers need to be aware of the range of feelings their students have during implementation (Harris & Brown, 2013) in order to be able to create a classroom climate of trust and respect, as this influences student participation in assessment practices (Brown, Andrade, & Chen, 2015). Affective threats are aggravated when teachers are not fully aware of students' emotional reactions to (peer) feedback (Higgins, Hartley, & Skelton, 2001). Positive teacher responses, such as showing empathy to students' emotions about (peer) assessment (Crossman, 2007) and insight into

students' emotional concerns, are therefore needed. To date, however, little is known about teachers' insight into the interpersonal dynamics of PA. This specific focus can be a starting point for the development of professionalization programs aimed at making teachers aware of these interpersonal dynamics. A first important step is to explore teachers' level of awareness regarding their students' perceptions of interpersonal processes and whether factors that were found to have a significant impact on students' perceptions of the interpersonal processes within PA can also be found when studying teachers' perceptions.

Peer Assessment: An Interpersonal Process

PA is fundamentally an interpersonal process as it generates thoughts, actions, motivational outcomes, and emotions for both assessees and assessors (Panadero, 2016). Attention to social and human factors is thus needed because well-implemented PA should decrease affective threats, assure accuracy, and lead to positive learning outcomes (Harris & Brown, 2016; Panadero & Brown, 2017; Topping, 2010).

In a recent survey study by Rotsaert et al. (2017), students' perceptions of the educational value of PA in relation to interpersonal variables, anonymity, and accuracy were investigated. In this study, and based on a recent review study by Panadero (2016), six frequently referred to interpersonal variables were explored: (1) *Friendship marking.* In this sense, friendship bonds have been identified as a source of potential scoring/feedback bias. However, only a small number of studies have directly addressed this topic (Panadero, Romero, & Strijbos, 2013). (2) *Fear of disapproval* refers to the assessors' fear of negative comments from the assessee if they give them a low score or critical feedback (i.e. recrimination) (Cartney, 2010). The results of a recent survey study on students' perceptions indicated that students' awareness levels regarding the fact that such processes can be present and influence the outcome of the PA activity exerted a positive effect on students' perceptions of the educational value of PA (Rotsaert et al., 2017). Furthermore, girls rated significantly higher on this factor than boys (Rotsaert et al., 2017). (3) *Psychological safety* refers to a situation in which students have a shared belief about taking interpersonal risks in a group. Creating a psychologically safe learning environment enables an environment in which differences of opinions are perceived as opportunities rather than conflicts (Nicol, 2010; Yu & Sung, 2015). (4) *Value congruency* refers to the importance of unanimity on both the goals and criteria of the PA activity (Cheng & Tsai, 2012). Rubrics hold the potential to augment value congruency within a PA activity as they provide the assessment criteria in a structured format and might thus enhance the perceived fairness and comfort with PA (Panadero et al., 2013). Rotsaert et al. (2017) found that both psychological safety and value congruency were not identified as significant predictors of students' perceptions regarding the educational value of PA. (5) *Trust in*

themselves as an assessor refers to the assessors' beliefs about their skills when assessing a peer (van Gennip et al., 2010). (6) *Trust in the other as an assessor* refers to the confidence in the reliability and validity of the assessment and feedback received from a peer. Students will only act on the basis of trustworthy information; if they believe that comments are capricious, they will not act on them (Carless, 2013). Trust in evaluative capabilities proved to be a significant predictor of students' perceptions of the educational value of PA (Rotsaert et al, 2017). Surprisingly, Panadero and Brown (2017) found that the constraints on PA implementation (e.g. negative class climate and distrust) reported by teachers did not significantly affect teachers' PA implementation.

Two factors are closely connected to the aforementioned interpersonal variables: the importance attributed to anonymity and perceived accuracy within PA. Regarding anonymity, Topping (1998) indicates that privacy is an important structural feature of PA in that disclosing the identity of the assessor or assessee seems to matter to the students. Vanderhoven et al. (2015) found that students have more positive attitudes toward PA when assessor anonymity was assured, while the participating teacher reported that revealing his/her identity worked as a means of controlling any undesirable social effects. Yu and Sung (2015) state that anonymity might offer more psychological safety for students, but at the same it might lead to misbehavior, such as positive marking towards friends. A recent survey study by Panadero and Brown (2017) reveals that the majority of participating teachers used an anonymous version of PA. The level of importance students' attributed to anonymity was found to be a significantly negative predictor for students' perception of the educational value of PA (Rotsaert et al, 2017). Furthermore, in this study, mean latent difference analyses indicated that the importance attributed to anonymity was lower for students with three or more than three PA experiences (Rotsaert et al, 2017).

Another crucial aspect related to interpersonal processes within PA is the concern about the (perceived) accuracy of all the actors in PA. Empirical research shows that students can be reliable sources under the appropriate conditions (Falchikov & Goldfinch, 2000; Topping, 2003), such as with the use of rubrics, involving students in the discussion of criteria, and/or considering the level of expertise of the students (for a detailed discussion, see Panadero et al., 2013). In the survey study on students' perceptions of PA, accuracy proved to be an important and positive predictor of the perceived educational value of PA (Rotsaert et al, 2017).

Assessment Policy in Flanders

As the deployment of assessment practices may be under the influence of national and/or regional assessment policies, it is important to become acquainted with the context in which this survey study was conducted. For decades, on both a meso and micro level, the Flemish government has

emphasized the importance of autonomy and trust in the policy-making capacity of schools. As part of this autonomy, teachers and teacher councils are, as a rule, solely responsible for the majority of pupil learning and classroom assessment. As a consequence, implementing formative assessment is the responsibility of individual teachers. Since 1998, the Flanders' educational government in Belgium has been urging educational institutions, by decree, to implement a competence model in teacher training programs. In the role descriptor *Teacher as guide of learning and developmental processes,* it is explicitly stated that the teacher needs to implement both assessment for and of learning. In a recent evaluation of Flemish teacher education centers, a majority of student teachers indicated that they were highly encouraged and trained to use new assessment methods, although teacher educators and mentors indicated the existence of a gap in the assessment culture between teacher educator centers (innovative) and the schools (conservative) in which student teachers conduct their internship. This gap was confirmed by the external quality agency, as their annual report stated that they were "worried about the alignment between educational goals, instruction methods, and assessment methods" (Vlaams Ministerie van Onderwijs en Vorming – Onderwijsinspectie, 2014, p. 4). In this study, we will focus on Flemish teachers' current application of PA in their daily classroom practice. Furthermore, it will be interesting to compare our results to the ones from the Spanish education setting studied by Panadero and Brown (2017).

Aim and Research Questions

The present study explores Flemish secondary school teachers' experiences with PA and seeks to obtain insight on a) teachers' level of awareness regarding students' considerations of the interpersonal processes within PA, and b) how teachers relate to their conceptions of the educational value of PA.

The research questions are:

RQ1– Do Flemish secondary school teachers report having experienced PA? If so, with what frequency and in what format?

RQ2a– Are teachers aware of students' perceptions on the interpersonal processes within PA?

RQ2b– What variables predict the educational value teachers attribute to PA?

Method

Participants

A total of 225 secondary school teachers participated in this study. 34.66% were teachers in grades 7-8 (henceforth *Level 1),* 34.66% were in grades 9-10 (henceforth *Level 2),* and 30.66% were in grades 11-12 (henceforth *Level 3).* The percentages of females and males were 64 % (N=144) and 36% (N=81) respectively. The distribution of the collected data over the four Flemish educational types is shown in Table 1.

Table 1
Distribution of the sample by type of school

Type of school	*N*	%
General Secondary	94	41.77
Technical Secondary	67	29.77
Arts Secondary	9	4.00
Vocational Secondary	55	24.44

Procedure

The survey conductors were bachelor's degree students in educational studies who were enrolled in a methodology course. In groups of four, they were asked to get permission from eight teachers to conduct a survey in their classes. In order to reduce social desirability, the pen-and-paper survey was designed to be filled in anonymously. Participants were also asked to complete an informed consent form, which contained information about the research purpose, confidentiality assurances, and the possibility of withdrawing. The consent forms were collected separately so as to ensure confidentiality. The survey conductors received a detailed coding manual (e.g. how to handle missing values), including a predefined Excel sheet to enter the data from the written surveys.

Instrument

A self-reporting survey instrument was designed with three blocks: (a) demographic information, (b) if applicable, descriptive questions about a PA activity respondents had organized in their class, and (c) specific questions about conceptions of PA and its social nature.

The items in block c of this questionnaire used a 6-point, positively packed rating scale in order to elicit more variance in responses (strongly disagree, mostly disagree, slightly agree, moderately agree, mostly agree, and extremely agree) (Brown, 2004). The following definition of PA was presented to ensure teachers, both with and without PA experience, had a shared understanding: "In a peer assessment-activity, students judge each other's tasks/presentations/ group assignments. This judgment can be expressed through scores, oral or written feedback, or a combination of both." Next, the content of blocks b and c will be explained in more detail (see Appendix 1):

a) PA activity description. The first eight questions established an overview of the kind of PA activity teachers had experienced, if any (block b).
b) Interpersonal processes in PA. The next six questions explored teachers' awareness with respect to students' concerns about the impact of interpersonal processes on PA outcomes (block c).
c) Importance of anonymity in PA. Three items explored the degree to which teachers thought students found anonymity important in PA (block c).
d) PA accuracy. Two questions were used to establish whether teachers thought PA was an accurate assessment method (block c).
e) PA conceptions. Five questions explored the value teachers' attributed to PA in terms of usefulness, involvement in the assessment process, and perceived learning gain (block c).

Analysis

To answer RQ1, we calculated the descriptive statistics from the data regarding teachers' previous PA experience and the PA format. Likewise, for RQ2a, after exploring the quality of the instrument and determining different factors through EFA and CFA analyses, descriptive statistics were calculated for the different scales. The goal of RQ2b was to understand how the educational value teachers attributed to PA was influenced by their awareness' level regarding students' perceptions of the interpersonal processes and the importance attributed towards anonymity and teachers' perceived accuracy of PA. Therefore, we identified the relationships between several latent factors and manifest item variables and explored their contribution to the perceived educational value of PA. For this, structural equation modeling (SEM) was used. By generating multiple parameters, SEM enabled us to incorporate the latent factors that were helpful in removing errors of measurement from the path analysis.

The following fit indices were calculated for every model: first, for *RMSEA*, values between 0 and .06 indicated a very good fit, and values between .06 and .08 indicated a reasonable fit.

Second, for *SRMR*, values between 0 and .08 indicated a very good fit. Third, for the *CFI* and *TLI* indices, acceptable values had to be larger than .90, and excellent values had to be above .95. Finally, the χ^2/df (chi squared/degrees of freedom) ratio was considered; for a value to be considered a good fit, it could not exceed 2.0 (Schreiber, Nora, Stage, Barlow & King, 2006; Schweizer, 2010).

Although our sample size was quite small, tests of measurement invariance (configural, metric, and scalar) were performed to explore the possible differences between educational levels, PA experience, and gender. To determine the measurement invariance across subgroups, it is preferable to report the change in *CFI* and *RMSEA* between the unrestricted and restricted models (Chen, 2007; Cheung & Rensvold, 2002). As Cheung and Rensvold (2002) recommend using a Δ*CFI* value higher than .01 to indicate a significant drop in fit and Chen (2007) suggests using Δ*RMSEA* to test for evidence of invariance, the criteria for invariance were Δ*CFI* ≤ .01 and Δ*RMSEA* ≤ .015. If measurement invariance is achieved, researchers can accept that different groups of individuals interpret the items and their underlying constructs in similar ways.

Results

RQ1–Do Flemish Secondary School Teachers Report Having Experienced PA, and in What Format?

As shown in Table 2, about one third of the teachers at all levels had not yet organized PA in their classrooms. Of the remaining teachers with experience, about 20 to 30 percent had more than three PA experiences.

Table 2
Teachers' experience with PA

Response category	Level 1 (*N*=71)		Level 2 (*N*=81)		Level 3 (*N*=73)	
	N	%	*N*	%	*N*	%
None	*26*	36.6	*36*	44.4	*22*	30.1
Once	*12*	16.9	*9*	11.1	*9*	12.3
Twice	*3*	4.2	*8*	9.9	*8*	11
Thrice	*9*	12.7	*6*	7.4	*5*	6.8
> Thrice	*21*	29.6	*22*	27.2	*29*	39.7

* Levels: Level 1 = Grade 7-8; Level 2 = Grade 9 -10; Level 3 = Grade 11-12

Looking at the specific features of the PA activities in a subsample of teachers with PA experience (Table 3), the results regarding teachers' consideration of whether to report the outcomes of the PA activity in the monthly report are equally spread over the three answer categories (yes/no/sometimes). Furthermore, this spread is seen at all levels. Together with the fact that most of the PA activities took place at the end of a series of lessons (between 55 and 65% of cases), it is possible that Flemish teachers currently see and apply PA more as a summative assessment method. Regarding student training in order to implement PA, almost half of the teachers at all three levels indicated not having trained their students to implement PA.

Table 3

Format of PA-usage for teachers with PA experience

Question and response category	Level 1 (*N*=41)		Level 2 (*N*=45)		Level 3 (*N*=49)	
	N	%	*N*	%	*N*	%
Grades: Was the result of the PA activity mentioned in the monthly report?						
Yes	*18*	43.90	*16*	35.56	*19*	38.77
No	*12*	29.27	*22*	48.89	*17*	34.70
Sometimes	*11*	26.83	*7*	15.56	*13*	26.53
Time: The PA activity took place						
…during a lesson	*18*	43.90	*15*	34.8	*19*	38.78
…at the end of series of lessons	*23*	56.10	*28*	65.12	*30*	61.22
Training: Were students trained to implement PA?						
Yes	*16*	39.02	*12*	26.66	*14*	28.57
No	*19*	46.34	*27*	60.00	*27*	55.10
Sometimes	*6*	14.63	*6*	13.33	*8*	16.33
Involvement in defining PA criteria: Were students involved in defining the PA criteria?						
Yes	*9*	21.95	*5*	11.11	*5*	10.20
No	*23*	56.10	*34*	75.55	*38*	77.55
Sometimes	*9*	21.95	*6*	13.33	*6*	12.24
Via what tool was the PA conducted?						
Paper	*19*	46.34	*18*	40.00	*29*	59.18
Technology (laptop/computer…)	*2*	4.88	*2*	4.44	*0*	0
Oral	*20*	48.78	*21*	46.66	*20*	40.82

As the provision of training to carry out PA is highly supported in current research guidelines, this result is worrying. Furthermore, the majority of teachers report not having involved their students in defining the PA assessment criteria, which is seen as an important part of training in PA guidelines. The PA mode was mostly paper-based or conducted orally. Surprisingly, as several in-class (e.g. electronic response systems) and online enablers have been developed in the last decade, technology-facilitated PA is almost totally absent.

Table 4
Anonymity within PA activity

Response category	Level 1 (*N*=41)		Level 2 (*N*=45)		Level 3 (*N*=49)	
	N	%	*N*	%	*N*	%
Anonymity for the assessor						
Yes	*11*	*26.83*	*19*	*42.22*	*11*	*22.45*
No	*21*	*51.22*	*22*	*48.88*	*29*	*59.18*
Sometimes	*9*	*21.95*	*4*	*8.88*	*9*	*18.36*
Anonymity toward the assessee						
Yes	*4*	*9.76*	*3*	*6.66*	*2*	*4.08*
No	*37*	*90.24*	*42*	*93.33*	*47*	*95.92*
Sometimes	*0*	*0*	*0*	*0*	*0*	*0*
Anonymity toward the teacher						
Yes	*0*	*0*	*0*	*0*	*0*	*0*
No	*41*	*100*	*45*	*100*	*49*	*100*
Sometimes	*0*	*0*	*0*	*0*	*0*	*0*

Table 4 provides details of teachers' responses regarding guaranteed anonymity within the experienced PA activities. The results show that teachers mostly decide not to create anonymous modes in PA. Only in some cases is anonymity guaranteed for the assessor.

RQ2a- Are Teachers Aware of Students' Perceptions on the Interpersonal Processes within PA?

Phase 1: Scale construction. To answer research questions 2A and 2B, first an exploratory principal axis factoring analysis was performed on a random 50% sample of the data. Following the recommendation of Hair et al. (2006), all items with loading of .50 or less were excluded from further analysis. Item(s) were also removed where the factor loading differed by .25 or less on

two factors. Such items were considered as having cross-loadings (Nunnally and Bernstein, 1994). Based on the first analysis, five items were deleted due to loadings across factors or low communality values. The second analysis was conducted on the remaining 11 items using a promax rotation, which allows factors to be correlated. Based on the scree plot, a four-factor solution was retained, which was also in line with the theoretical model. Table 5 shows the factor loadings after rotation. The Kaiser–Meyer–Olkin measure verified the sampling adequacy for the analysis, KMO = .77, which is above the commonly recommended value of .6, and Bartlett's test of sphericity indicated $\chi^2(55) = 509.73$, $p < .001$.

Table 5
Results of the exploratory factor analysis (EFA)

		F1	F2	F3	F4
1	I'm aware of the fact that pupils might have insufficient trust in each other's' evaluative capabilities.	**.867**	.052	.132	-.132
2	I'm aware of the fact that pupils might have fear of the possible consequences of giving a low score or negative feedback.	**.844**	-.122	-.022	.055
3	I'm aware of the fact that pupils might lack insight or disagree on the different criteria being used.	**.752**	.138	.058	-.083
4	I'm aware of the fact that pupils might have insufficient trust in their own evaluative capabilities.	**.647**	-.139	-.034	.042
5	I'm aware of the fact that pupils might not find the class atmosphere safe enough to assess or give feedback to their peers.	**.575**	.103	-.219	.183
6	I think students find it important that the identity of the assessee is hidden.	-.076	**.963**	.020	-.071
7	I think students find it important that the assessee doesn't know the identity of the assessor*. * The anonymity of the assessor is guaranteed.	.030	**.901**	-.048	.087
8	Evaluation through PA is accurate.	.007	.056	**.866**	.020
9	Students are capable of giving an accurate judgment to each other.	-.056	-.127	**.595**	.118
10	Through participation in PA activities, students feel actively involved in the assessment process.	.042	-.101	-.023	**.718**
11	It is instructive for pupils to give and receive scores and feedback to/from peers	-.003	.143	.171	**.664**

Pattern matrix, promax rotation. The number in bold represents the factor loading linked to each factor, respectively, F1, F2, F3, and F4.

The first factor is "interpersonal factors within PA", referring to the extent to which teachers have considered the effect of interpersonal processes on the PA process from a student's point of

view. The second factor is the "importance of anonymity," which evaluates teachers' awareness about the importance students attribute to guaranteeing anonymity for assessors and/or assessees. Anonymity towards the teacher could not be identified as one of the observed items contributing to this factor. The third factor is "accuracy," which measures teachers' perceptions of the accuracy of PA and whether they think students are capable of giving accurate judgments. The fourth factor is the "educational value of PA", referring to the extent to which teachers think PA is a valuable assessment method. Based on the structure found in the EFA, we fitted a measurement model (i.e. confirmatory factor analysis). The results show an acceptable fit between the hypothesized model and the observed data $\chi^2(38) = 71.72$, $\chi^2/df = 1.88$, $p < .001$, CFI = .958, TLI = .939, SRMR = .053 and RMSEA= .063, with a 90% confidence interval between .040 and .085.

Phase 2: Reliability analysis and descriptive statistics. Since our instrument taps several conceptually different constructs, each of them contributing to an overall formative index (Educational value of PA), it is not recommended to evaluate the instrument's reliability through one internal consistency coefficient such as Cronbach's alpha (Sijtsma, 2009; Streiner, 2003). Instead, we assessed the reliability of our measures by computing the omega coefficient for each of the latent factors. This strategy is often recommended for assessing reliability when a latent factor structure is available (McDonald, 1999). The omega value for Interpersonal processes within PA was $\omega = 0.98$, for Importance of anonymity $\omega = 0.71$, for Accuracy $\omega = 0.67$, and for Educational value of PA $\omega = 0.85$, showing high to acceptable reliability. Table 6 presents the mean scores (M) and standard deviation (S.D.) for each factor, ranging from a minimum score of 1 to a maximum of 6. Results were moderately positive for the first three factors and highly positive for the educational value of PA.

Table 6
Omega value and descriptive statistics

	ω	M	S.D.
Interpersonal factors within PA	.98	3.27	1.02
Importance of anonymity	.71	3.03	1.37
Accuracy	.67	3.47	.86
Educational value of PA	.85	4.70	.86

The correlations between the three latent predictors, shown in Table 7, confirm that these factors are related to one another.

Table 7
Correlation between latent predictors

$p < 0.001$	Importance of anonymity	Accuracy
Interpersonal factors within PA	.403	-.334
Importance of anonymity		-.278*

* p =.049

RQ2b – What Variables Predict the Educational Value Teachers Attribute to PA?

In order to correctly identify the model, the variance of all four latent variables was fixed to one. Since we wanted to know what variables predict the "*educational value of PA*", we forced this latent construct to be explained (i.e. receive loadings) by the other three. The full SEM model (with standardized loadings) is shown in Figure 1. The results show an acceptable fit between the hypothesized model and the observed data ($\chi 2=329.6$, df=40, $\chi 2/df= 1.85$, p =0.007409). The goodness of fit estimates were CFI = .958, TLI = .942, SRMR = .053 and RMSEA= .062, with a 90 % interval of .039 and .083. The results suggest that all items load significantly onto the four latent factors. The item loadings range between .49 and .93.

As can be seen in Figure 1, only *accuracy* ($\beta = .565$, *std. error* = .124, $p < .001$) proved to have a significant impact on teachers' *educational value of PA* ($R^2 = .319$ (*std. error* = .141). Furthermore, the SEM confirms that these variables are connected to one another.

Although we had a relatively small sample, tests of measurement invariance (configural, metric and scalar) were performed to explore the possible differences between PA experience, educational levels, and gender. However, the changes in the CFI and RMSEA values between the unrestricted and restricted models were higher than .01 and .015 respectively. As a consequence, tests on differences in the latent means were not performed.

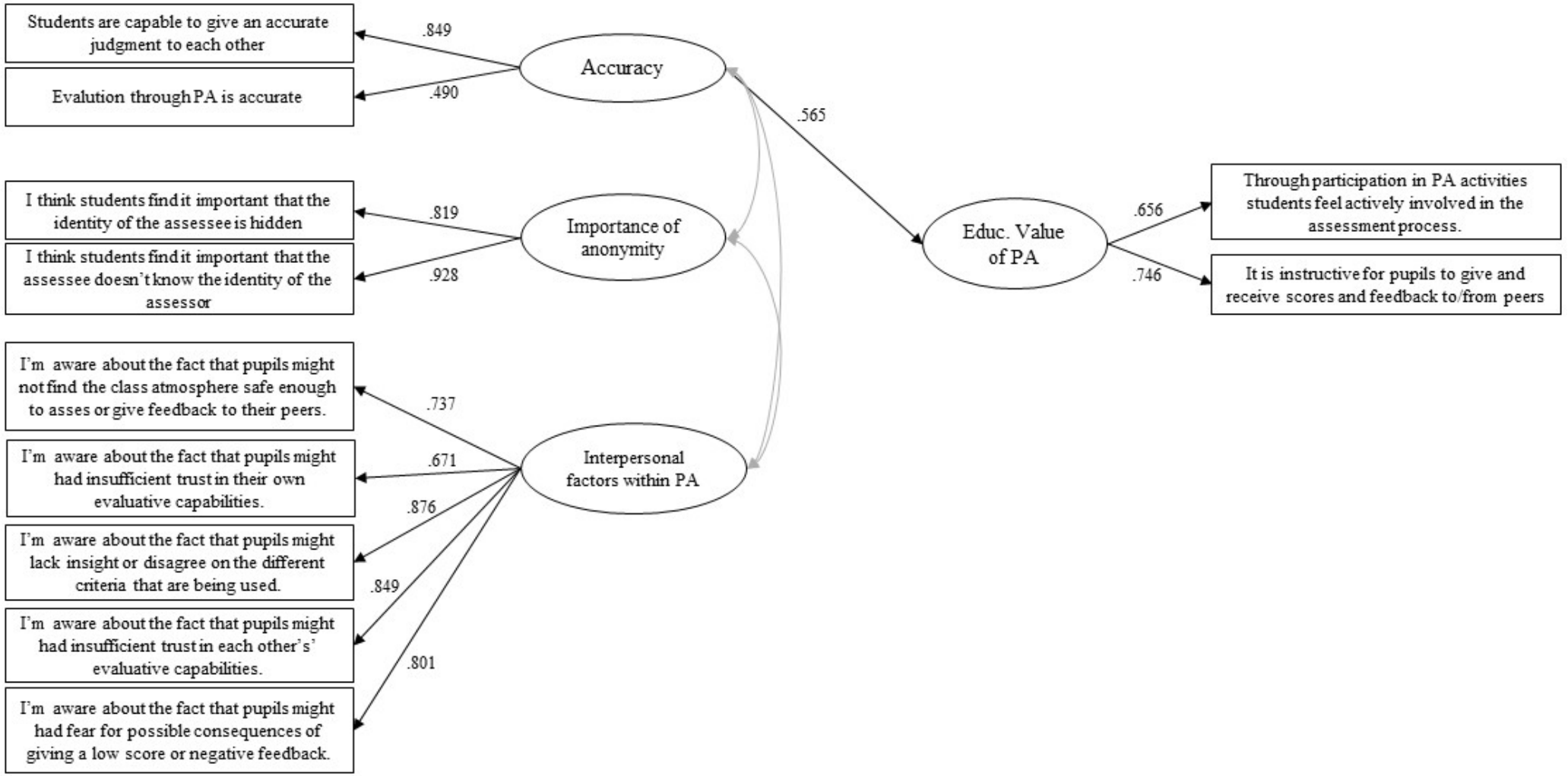


Figure 1 *Baseline SEM model with standardized loadings.*

Discussion

The present study examined Flemish secondary school teachers' experiences with PA as well as the teachers' insights on the social nature of PA. The interpersonal nature of PA and peer feedback is an under-explored area of research. Additionally, with a few exceptions, research has tended to focus on students' perceptions. However, a successful implementation of PA practices relies on the teacher's ability and motivation to adequately prepare students for this activity (Harris & Brown, 2013; Panadero & Brown, 2017). A part of this preparation involves teachers' understanding of students' interpersonal concerns that may arise as a consequence of their participation in PA. To date, however, little is known about teachers' insights into the interpersonal dynamics of PA, which is a new line of research opened up in this study.

The results of RQ1 on teachers' current use of PA show that the majority (63%) of the 225 teachers had previous experience with PA, of which about one-third at all three levels had multiple experiences. It is noteworthy that the reported frequency of PA use is moderately higher than that found in several previous studies. For example, in Noonan and Duncan (2005), 49% of teachers reported little use of PA and self-assessment whereas this figure was 55% in the study of Panadero & Brown (2017). The results indicate that PA is part of Flemish teachers' armamentarium of innovative assessment methods. Furthermore, the fact that about one-third of the teachers had multiple PA experiences is consistent with previous findings indicating that teachers' self-reported use largely depends on their previous positive experiences (Panadero & Brown, 2017). However, the low levels of both training and involvement of students when defining the PA criteria suggest that the assessment activities are generally superficially embedded within the learning environment. The plausibility of this interpretation is strengthened by the finding that most of the PA activities at all educational levels took place at the end of a series of lessons, highlighting what could be a rather summative approach to PA. Furthermore, in contrast to the Spanish sample of teachers in Panadero and Brown's (2017) study, the majority of this group of teachers used non-anonymous modes of PA. This might indicate that the teachers in our sample are unaware that students might be resistant to participate in PA. Nevertheless, there is still a tension to be resolved between anonymity and formative PA use (Panadero, 2016). Regarding RQ2a, results show that teachers in the studied population have a moderate awareness of students' concerns regarding the interpersonal processes in PA as well as the importance students attribute to anonymity within PA. This indicates a possible lack of insight and knowledge about these processes. Furthermore, PA accuracy is moderately positively evaluated by teachers. Overall, the educational value of PA is highly positively appraised. This lack of teacher awareness of students' concerns for PA interpersonal processes might lead to an aggravation of affective

threats, such as discomfort and fear of disapproval and/or distrust (Higgins, Hartley, & Skelton, 2001). It is for further research to explore how to raise teachers' considerations for the human and social nature of these practices as well as how to create a trusting and respectful PA learning environment (Brown & Harris, 2016). A valuable intervention in this regard might be to primarily focus on teachers' general conceptions of assessment before focusing on the training of PA literacy (i.e. understanding and use of peer assessment) (Xu & Brown, 2016). Suitable assessment conceptions (i.e. assessment is for learning) are an essential precondition because conceptions denote the belief systems that teachers have about assessment and enclose their cognitive and affective responses to specific assessment practices (Boud, 2016; Xu & Brown, 2016). The practical realization of such an approach will be discussed in the implications section.

In order to understand the relationship between the identified latent constructs of teachers' awareness of influencing interpersonal processes, anonymity, perceived accuracy and the value teachers attribute to PA (RQ2b), SEM was used. As this is the first investigation of such an under-researched issue, the presented model should be seen as a first step in disentangling the complexity of the studied topic. The results lead to the conclusion that in this sample 31.9% of the variance in the *Educational value of PA* can be solely explained by teachers' perceived accuracy of PA. The accuracy theme within PA research has been of major concern for both researchers and practitioners. However, the fact that accuracy was the only identifiable predicting factor raises concern. That is, this finding might imply that, in teachers' minds, the goal of PA is predominantly aimed at scoring instead of creating a two-way dialogic peer feedback environment, as was also reflected in the descriptive data (RQ1). No other latent factors were found to predict teachers' perception of the educational value of PA. Although perceived accuracy was also found as an important predictor of students' perception of the educational value of PA in our survey study (Rotsaert et al., 2017), the discrepancy between students' perceptions towards interpersonal variables and related aspects (more specifically negative interpersonal variables, trust in evaluative capabilities, and the importance attributed to anonymity) and teachers' awareness level regarding these students' perceptions is worrying. Again, this finding has important consequences for teacher training and professional development in assessment literacy (see further).

An important limitation of this study is its self-reported nature. As this study only uses a measure of teachers' awareness levels regarding interpersonal processes, future studies could triangulate teachers' embraced beliefs with their enacted behavior in the classroom. Also, teachers' answers may have been influenced by social desirability, which is a risk with any form of subjective data collection (Desimone, 2009). However, throughout the process of survey development and administration, several steps were taken to reduce social desirability bias,

including critical reviews and pre-testing by an expert in formative assessment. Moreover, confidentiality for respondents was assured.

It might be that some important aspects of the latent constructs were not adequately highlighted by the questions in our instrument. Our study involved complex psychological constructs, and the model included a limited number of items for measuring each one. This is in line with previous research showing that multifaceted psychological constructs can be measured through short scales (Rammstedt & John, 2007). In any case, it would be interesting to include more measures of these same constructs in future studies.

Implications and Future Lines of Research

Two key implications for both practice and research are worth summarizing. First of all, based on the finding that teachers' perceived accuracy of PA is of significant importance for their beliefs in the educational value of PA, it becomes apparent that in teachers' minds, PA should currently be used as a summative grading method rather than as a formative embedded feedback practice. This is in contrast with the move the research on assessment for learning methods has gone through in the past decade, in which the distinction between assessment and instruction has been softened. In this view, (peer) feedback is seen as a process of two-way communication and dialogue, rather than a one-way transmission of information from teacher to student, which positions students in the center of the PA activity as active learners (Nicol, Thomson, & Breslin, 2014; Sadler, 2010; Yang & Carless, 2013). This new way of thinking has been formulated as a shift in focus from providing feedback to the embedding of feedback within the learning design and emphasizing interactions between students and teachers (Boud & Molloy, 2013). This implies that PA should no longer be seen as an episodic mechanism (Rowe, 2017), requiring unsophisticated PA practices to be left behind (Panadero, 2016). As a practical consequence, this calls for PA practices in which teachers: a) clarify the purpose of PA, its rationale, and expectations to the students; b) involve students in developing and clarifying assessment criteria; c) match participants in a way that fosters productive PA; d) provide quality PA training, examples, and practice; e) provide rubrics, scripts, checklists, or other tangible scaffoldings for PA; f) specify PA activities and timescales; and g) monitor the PA process and coach students (Panadero, Jonsson, & Strijbos, 2016, p. 10). In this sense, involvement in PA has the potential to promote pro-social behavior, calm emotions, and contribute to a sense of belonging (Rowe, 2017).

Secondly, our findings have important implications for teachers and teacher educators' professional development. In general, professional development activities are an obvious starting point for enhancing teachers' (peer) assessment literacy (i.e. understanding and use of assessment and related skills to apply the knowledge that comes from assessment) (Deneen &

Brown, 2016), for example, through communities of practice in which 'accounts of practice' surface and are shared through the process of thoughtful reflection on practice (Hounsell & Zou, 2017). With regard to this study, in these communities of practice recommendations made by empirical researchers in authentic settings that are directed to the teacher audience (e.g. the possible positive effects of the use of anonymity in face-to-face classroom settings or insights on interpersonal processes within PA) can be discussed and reflected upon. These accounts of practice seem to have a high degree of 'street credibility' for fellow teachers as they exemplify how compromises can be made between contextual factors external to assessment (e.g. class size and teaching schedule), student needs, and teachers' perceptions.

Conclusions

This study adds to our understanding of teachers' PA use and opens up a new venue for research on teachers' awareness of PA interpersonal processes in their students. Furthermore, the relationship between teachers' awareness of these processes and the educational value teachers attribute to PA was examined. Our results from the Flanders context, a high self-governance school policy context, show that teachers have a moderate awareness of the interpersonal processes in PA as well as the importance their students attribute to anonymity within PA activities. However, there is certainly room for improvement regarding teachers' insights and mastery of these interpersonal processes. More specifically, this study points out that teachers' perceived accuracy is a major predictor of their belief in the educational value of PA. In this sense, this study confirms previous work on teachers' conceptions of PA, in which shallow, unsophisticated PA approaches were found. For teachers to implement PA as a learning oriented assessment method, both pre-and-in-service professional development initiatives will have to offer thoroughgoing insights into both their own and students' general and method-specific assessment perceptions, next to guidance on the operationalization of existing practice recommendations made by empirical researchers.

References

Bolzer, M., Strijbos, J. W., & Fischer, F. (2015). Inferring mindful cognitive-processing of peer-feedback via eye-tracking: role of feedback-characteristics, fixation-durations and transitions. *Journal of Computer Assisted Learning*, *31*(5), 422–434. doi:10.1111/jcal.12091

Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151–167. doi:10.1080/713695728

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, *38*(6), 698–712. doi:10.1080/02602938.2012.691462

Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 1–14. doi:10.1080/02602938.2015.1018133

Brown, G. T. L. (2004). Measuring attitude with positively packed self-report ratings: comparison of agreement and frequency scales. *Psychological Reports*, *94*(3), 1015–1024. doi:10.2466/pr0.94.3.1015-1024

Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, *22*(4), 444–457. doi:10.1080/0969594X.2014.996523

Brown, G. T. L., & Harris, L. R. (2016). *Handbook of Human and Social Conditions in Assessment*. New York, NY: Routledge.

Carless, D. (2013). Trust and its role in facilitating dialogic feedback. In *Feedback in Higher and Professional Education* (pp. 90–103). doi:10.4324/9780203074336

Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education*, *35*(5), 551–564. doi:10.1080/02602931003632381

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255.

Cowie, B., & Harrison, C. (2016). Classroom Processes that Support Effective Assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 335–350). New York, NY: Routledge.

Crossman, J. (2007). The role of relationships and emotions in student perceptions of learning and assessment. *Higher Education Research & Development*, *26*(3), 313–327. doi:10.1080/07294360701494328

Deneen, C. C., & Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, *3*(1), 1225380. doi:10.1080/2331186X.2016.1225380

Desimone, L. M. (2009). Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher*, *38*(3), 181–199. doi:10.3102/0013189X08331140

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education*, *24*, 331–350. doi:10.1080/03075079912331379935

Falchikov, N. (2003). Involving students in assessment. *Psychology Learning & Teaching*, *3*, 102–108.

Falchikov, N., & Goldfinch, J. (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, *70*(3), 287–322. doi:10.3102/00346543070003287

Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). Multivariate data analysis (6th ed.).Uppersaddle River, N.J.: Pearson Prentice Hall

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, *36*, 101–111. doi:10.1016 /j.tate.2013.07.008

Harris, L. R., & Brown, G. T. L. (2016). Volume Introduction: The human and social experience of assessment: Valuing the person and context. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 1–17). New York, NY: Routledge.

Higgins, R., Hartley, P., & Skelton, A. (2001). Getting the Message Across: The problem of communicating assessment feedback. *Teaching in Higher Education*, *6*(2), 269–274. doi:10.1080/13562510120045230

Hounsell, D., & Zou, T. X. P. (2017). Surfacing and Sharing Advances in Assessment: A Communities-of-Practice Approach (pp. 33–48). Springer Singapore. doi:10.1007/978-981-10-3045-1_3

Nicol, D. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, *35*(5), 501–517. doi:10.1080/02602931003786559

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102–122. doi:10.1080/02602938.2013.795518

Noonan, B., & Duncan, C. R. (2005). Peer and Self-Assessment in High Schools. *Practical Assessment, Research and Evaluation*, *10*(17), 1–8.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric Theory (3rd ed.). New York: McGraw-Hill.

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 247–266). New York, NY: Routledge.

Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: positive experience predicts use. *European Journal of Psychology of Education*, *32*(1), 133–156. doi:10.1007/s10212-015-0282-5

Panadero, E., Jonsson, A., & Strijbos, J.-W. (2016). Scaffolding Self-Regulated Learning Through Self-Assessment and Peer Assessment: Guidelines for Classroom Implementation. In *Assessment for Learning: Meeting the Challenge of Implementation* (pp. 311–326). doi:10.1007/978-3-319-39211-0

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, *39*(4), 195–203. doi:10.1016/j.stueduc.2013.10.005

Raes, A., Vanderhoven, E., & Schellens, T. (2013). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Studies in Higher Education*, *0*(0), 1–16. doi:10.1080/03075079.2013.823930

Reinholz, D. (2015). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 1–15. doi:10.1080/ 02602938.2015.1008982

Rotsaert, T., Panadero, E., Estrada, E., & Schellens, T. (2017). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. *Studies in Educational Evaluation*, 53, 29-40. doi: 10.1016/j.stueduc.2017.02.003

Rowe, A. D. (2017). Feelings About Feedback: The Role of Emotions in Assessment for Learning (pp. 159–172). Springer Singapore. doi:10.1007/978-981-10-3045-1_11

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, *35*, 535–550.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research, 99*, 323-338. doi:10.3200/JOER.99.6.323-338

Schweizer, K. (2010). Some Guidelines Concerning the Modeling of Traits and Abilities in Test Construction. *European Journal of Psychological Assessment, 26*, 1-2. doi:10.1027/1015-5759/a000001

Streiner, D. L. (2003). Being Inconsistent About Consistency: When Coefficient Alpha Does and Doesn't Matter. *Journal of Personality Assessment, 80*, 217-222. doi:10.1207/S15327752JPA8003_01

Topping, K. J. (1998). Peer Assessment between Students in Colleges and Universities. *Review of Educational Research*, *68*(3), 249. doi:10.2307/1170598

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), (pp. 55–87). Dordrecht, The Netherlands: Kluwer Academic.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*(4), 339–343. doi:10.1016/ j.learninstruc.2009.08.003

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, *20*, 280–290. doi:10.1016/j.learninstruc.2009.08.010

Vlaams Ministerie van Onderwijs en Vorming – Onderwijsinspectie (2014) *Onderwijsspiegel.* Retrieved from http://www.onderwijsinspectie.be/sites/default/files/atoms/files/ Onderwijsspiegel_2014.pdf

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, *18*(3), 1–13. doi:10.1080/13562517.2012.719154

Yu, F.-Y., & Sung, S. (2015). A mixed methods approach to the assessor's targeting behavior during online peer assessment: effects of anonymity and underlying reasons. *Interactive Learning Environments*, 1–18. doi:10.1080/10494820.2015.1041405

Appendix 1

Teachers' conceptions of PA

Definition: In a PA activity, students judge each other's tasks/presentations/group assignments. This judgment can be expressed through scores, oral or written feedback, or a combination of both.

#	Question	Response Format
1	Have you ever organized a PA-activity in your class?	Once –Twice –Thrice – > Thrice – None
2	Via what tool was the PA conducted?	Paper – Oral – Computer – Combination
3	Were the results of the PA accounted for in the monthly report?	Y/N/Sometimes
4	At what moment did the PA activity take place?	During / At the end of a series of lessons
5	Anonymity was guaranteed for: a) Assessors b) Assessees c) Teachers	Y/N/Sometimes in each of three categories
6	Were students trained to do PA (e.g. through a teacher example or an instruction video)?	Y/N/Sometimes
7	Were students involved in defining the assessment criteria?	Y/N/Sometimes
8	Was the result of the PA orally discussed with you?	Y/N/Sometimes
9i	Interpersonal - friendship marking: I'm aware of the fact that students might think that PA results were influenced by mutual friends giving each other inaccurate judgments and feedback.	6 point Likert-scale (2 negatively formulated, 4 positively formulated)
9ii	Interpersonal - fear of disapproval: I'm aware of the fact that pupils might have fear of the possible consequences of giving a low score or negative feedback.	Idem

9iii	Interpersonal - trust peer: I'm aware of the fact that pupils might have insufficient trust in each other's' evaluative capabilities.	Idem
9iv	Interpersonal - trust own: I'm aware of the fact that pupils might have insufficient trust in their own evaluative capabilities.	Idem
9v	Interpersonal - psychological safety: I'm aware of the fact that pupils might find the class atmosphere not safe enough to assess or give feedback to their peers.	Idem
9vi	Interpersonal - value congruency: I'm aware of the fact that pupils might lack insight or disagree on the different criteria being used.	Idem
10i	Anonymity: I think students find it important that the assessee doesn't know the identity of the assessor*. * The anonymity of the assessor is guaranteed.	Idem
10ii	Anonymity: I think students find it important that the identity of the assessee is hidden.	Idem
10iii	Anonymity: I think students find it important that the identity of the assessor is hidden to the teacher.	Idem
11i	Accuracy: Students are capable of giving an accurate judgment to each other.	Idem
11ii	Accuracy: Evaluation through PA is accurate.	Idem
12i	PA conceptions: Through participation in PA, students get to know their strengths and weaknesses.	Idem
12ii	PA conceptions: PA is useful.	Idem
12iii	PA conceptions: It is instructive for pupils to give scores and feedback to peers.	Idem
12iv	PA conceptions: Through participation in PA activities, students feel themselves responsible for the assessment process.	Idem
12v	PA conceptions: Through participation in PA activities, students feel actively involved in the assessment process.	Idem

4 Using Mobile Response Technology for anonymous PA and PF: the interrelatedness of interpersonal variables and students' preferred type of teacher assessment

This chapter is based on:

Rotsaert, T., Schellens, T., Raes, A., & Van Hoey, S. (submitted). Use of Mobile Response Technology for Anonymous Peer Assessment and Feedback: Exploring the Interrelatedness of Interpersonal Variables and Students' Preferred Type of Teacher Assessment. Manuscript under review for publication in *Journal of Educational Computing Research*

Chapter 4
Using Mobile Response Technology for anonymous PA and PF: the interrelatedness of interpersonal variables and students' preferred type of teacher assessment

Abstract

The aim of this study is to explore the use of Mobile Response Technology (MRT) as a facilitator for immediate anonymous peer assessment (PA) and peer feedback (PF) in face-to-face classroom settings. In this study 39 university students' perceptions regarding MRT, specific interpersonal variables within PA and preferred type of teacher assessment were studied. Results show that students highly appreciate the anonymous setting that was created by means of MRT and the features it can offer. This study found that students' conceptions of peer assessment were significantly related to students' fear of disapproval when giving low scores or negative feedback, the amount of congruency regarding the goals and the criteria of the assessment activity, and students' trust in their peers' evaluative capabilities. Finally, students expressed their preference for non-anonymous teacher feedback and assessment. Implications for future research are discussed.

Introduction

Students' involvement in assessment processes contributes to their empowerment, and should provide them with the necessary assessment skills for their professional development and lifelong learning (Planas Lladó et al., 2014). Peer assessment (PA) is an example of a collaborative assessment method in which students are actively involved. In PA activities students are required to reflect on the quality of a peer's work and discuss how well it corresponds with the explicitly stated criteria of the work (Strijbos & Sluijsmans, 2010). Studies on PA indicated that positive results can only be achieved when certain preconditions are met, such as training in PA (Liu & Li, 2013; Sluijsmans, Brand-Gruwel, van Merriënboer, & Martens, 2004; Van Zundert, Sluijsmans, Konings, & Van Merrienboer, 2012), and the use of rubrics (Panadero, Romero, & Strijbos, 2013). Despite all of the work that has been accomplished, research on PA is still in its infancy (Kollar & Fischer, 2010;Panadero, 2016). Only recently, research has focused on the inherently social nature of PA and the possible reciprocity effects caused by interpersonal processes such as friendship marking, psychological unsafety, fear of disapproval when giving a low score or

negative feedback, distrust in own and others' evaluative capabilities, and value diversity on the goals and criteria of the activity (Panadero, 2016; van Gennip, Segers, & Tillema, 2010; Vanderhoven, Raes, Montrieux, Rotsaert, & Schellens, 2015). More specifically, van Gennip et al. (2010) found in a study in vocational secondary education that interpersonal variables play a significant role in peer assessment settings and are of significant influence on students' conceptions of PA and perceived learning. Additionally, previous research has shown that providing anonymity for the assessors can help to relieve the interpersonal burden from students (Yu & Liu, 2009) and can lead to more positive conceptions towards PA (Vanderhoven et al., 2015). Therefore, building on van Gennip and colleagues' model, we investigate interactional processes in PA within an anonymous PA-setting. As providing anonymity in real-time face-to-face situations is a challenge, a possible intervention in tackling this issue is the implementation of clickers (i.e. polling systems), which - through its provision of anonymity for the assessors - has proven to be an adequate facilitator in reducing undesirable social effects within face-to-face peer assessment settings (Vanderhoven et al., 2015). Despite the aforementioned advantages, students criticised the lack of immediate feedback which clickers were unable to offer (i.e. only scores can be generated with clickers). In this study, Mobile Response Technology will be investigated as a possible solution for facilitating both immediate anonymous peer assessment and peer feedback in the classroom, as this web based application has a feature to give both anonymous scores based on rubrics as well as anonymous feedback.

Next to the enabling option of immediate anonymous assessment and feedback, the use of MRT in face-to-face PA settings also redefines the role of the teacher within assessment. That is, teacher enabling processes will be needed in order to enhance students' skill development of generating PA scores and peer feedback (PF) during the provision, moderation and reception of PA and PF (Xu & Carless, 2016). In sum, this study aims to contribute to our understanding of a) the use of MRT in face-to-face PA environments, b) the interrelatedness of interpersonal variables and their relations with conceptions of peer assessment in an anonymous peer assessment and feedback setting, and c) students' preferred type of teacher assessment within this settings.

Mobile Response Technology and Peer Assessment

Mobile devices such as laptops, tablets, and smartphones have become a learning tool with great potential (Sung, Chang, & Liu, 2015). With regard to classroom response technology, the prevalence of portable technology is welcome as it eliminates the high cost of classroom technology hardware (Güler, 2016), at least when students are able and/or allowed to bring

their own device and if free web-based services are used. Mobile Response Technology (MRT) has been put forward as a valuable answer to this issue. MRT operates via a free web-based application on web-enabled devices via wi-fi or mobile internet (i.e. smartphone, tablet, laptops and desktops) wherein students can give scores and written feedback to each other. Magaña and Marzano (2014) indicate that nonetheless they have been developed for several years now, educational researchers have not yet thoroughly studied their applicability in peer assessment and feedback settings. Frequently used software applications are PollEveryWhere, Mentimeter, Socrative and WhatsApp (Güler, 2016). The benefits of MRT include having students see the distribution of class responses when projected live in real-time in the classroom and mobile devices are suitable for sending responses to open-ended questions by instant messaging (Stowell, 2015). In this study we chose to use the MRT application Socrative. Conversely, mobile devices introduce the possibility of digital distraction, frustration with unreliable internet connections, and possibly added cost to the student. Merely incorporating response technology into a course does not mean more learning is guaranteed; rather, its effectiveness as a learning tool is activated when they are used as a component of a constructivist, student-centered pedagogy (White, Syncox, & Alters, 2011). Previous studies (e.g., Caldwell, 2007) suggest that response technology may have positive impacts on teaching and learning in education. In terms of feedback practices with response technology, they are currently generally used as an interaction tool in which teachers present statements on learning content in combination with a peer discussion (Hunsu, Adesope, & Bayly, 2016).

As these new technologies can capture and easily share both student and teacher input, they may also have the potential to facilitate (peer) assessment and feedback processes (Magaña & Marzano, 2014). Assessment research emphasizes that students require practice and training to become skilled peer assessors and assessees, who provide and receive high-quality PF (Sluijsmans & Brand, 2002). Teachers are thus challenged to implement assessment activities in which students are prompted to provide frequent PF, resulting in a frequent enactment of the peer-assessor role (Tsivitanidou & Constantinou, 2016). As briefly mentioned in the introduction, the study of Raes et al. (2013) found that when introducing anonymous peer assessment using clickers, students endorsed the added written and oral peer feedback, which provided them with the needed argumentation and justification (i.e. the qualitative feedback part) of the given scores on rating scales (i.e. the quantitative assessment part). For the score part in this setting immediate visual feedback of the scores was guaranteed. However, this was not the case for the written feedback part which was even more preferred and was processed by the students at home one day after the peer assessment session. Earlier studies showed that when writing feedback, students

have more opportunities to engage in cognitive activities such as critical thinking, planning, monitoring, and regulation (Lin, Liu, & Yuan, 2001). This lack of immediate written feedback of the assessment activity hinders the creation of a rich feedback and learning environment, where assessment and feedback practices are not viewed as 'added on' activities (Havnes, Smith, Dysthe, & Ludvigsen, 2012). As stated by Nicol (2010) it would add more 'pedagogical power' to a learning environment when it provides both immediate (peer) assessment and feedback, which provides information on how to improve future performance and learning.

Anonymity and interpersonal processes in PA

The interplay of interpersonal variables is inherent to PA-practices. As this interplay might influence the assessment outcome, it has often been stated that decreasing negative social effects via the introduction of anonymity is desirable (Ballantyne, Hughes, & Mylonas, 2002; Vickerman, 2009; Yang & Tsai, 2010; Yu & Sung, 2015), or should at least be explored (Howard, Barrett, & Frick, 2010). Attention to these factors is thus needed because well implemented PA should decrease negative social problems and lead to positive learning outcomes (Panadero, 2016; Topping, 2010). Consequently, it has been suggested that the relevance of anonymity should be studied in synchronous educational settings (Ainsworth et al., 2011). A useful concept for studying the acceptance of the PA-practice is that of its perceived usefulness and the way in which students concur with the intended learning objectives of the assessment practice (van Gennip et al., 2010). These two concepts are taken along as dependent variables in this study. The difference between the two selected dependent variables lies in the fact that *general conceptions* (CON PA) represent students' perceptions of PA (i.e. whether it is perceived as useful), while *conceptions of perceived learning* (CONL PA) focuses on students' considerations of the learning objectives in PA (i.e. the perceived added learning value of being involved in these processes) (Cheng & Tsai, 2012). Building on the latest findings from the field, six interpersonal variables were chosen because they are the most relevant when it comes to their possible effects on the perceived educational value of PA (CONL PA): *value congruency, psychological safety, friendship marking, fear of disapproval and trust in own and others' evaluative capabilities.*

Value congruency (VCG) on the goals and criteria of the PA activity: van Gennip et al. (2010) were the first to include a measure of unanimity of goals in research on the social nature of PA, as it is an important predictor for a meaningful assessment process. Remarkably, van Gennip et al. (2010) found that value congruency on goals was a negative predictor of conceptions towards PA (CON PA), but was in turn positively related to

perceived learning (CONL PA). Considering the importance of shared criteria, in their qualitative analysis, Cheng and Tsai (2012) elaborate on this variable and argue that it should be investigated for both the selected goals and the assessment criteria of the PA-activity. Therefore, both are included in our definition of this variable. *Psychological safety* (PSY), refers to a situation in which students have a shared belief about taking interpersonal risks in a group. People that feel psychologically safe tend to perceive differences in opinions as opportunities rather than conflicts (Yu & Sung, 2015). Several authors state that creating a safe environment is a precondition for accurate and thus valuable PA activities (Harris & Brown, 2013; van Gennip et al., 2009). *Friendship marking* (FR) due to friendship bonds has been frequently mentioned as a source of potential scoring/feedback bias. However, only a small number of studies have directly addressed this topic (Strijbos et al., 2009). Recent research on the diminishing effect of rubrics on over- and underscoring (i.e. friendship marking) by peers in PA shows that for low and medium friendship levels, a rubric does reduce the friendship bias, but for high-level friendship, the rubric even seems to amplify the potential friendship bias (Panadero et al., 2013). Cheng and Tsai (2012) found that anonymity was preferred for the reason of avoiding the pressure of friendships. *Fear of disapproval* (FDP) when giving a low score or critical feedback refers to the assessors' fear of negative comments from the assessee if they give a low score or critical feedback (i.e. recrimination) (Cartney, 2010; Lu & Bol, 2007). *Trust in own* (TRS) *and others' evaluative capabilities* (TRP) appears to be important for PA-practices, as it is an essential factor for both the assessors (i.e. feeling capable of giving valuable assessment and feedback) and assessees (i.e. willing to receive and act upon the given assessment and feedback in subsequent performances) (Carless, 2013).

Role of the teacher within PA

Harris and Brown (2013) suggested that a successful implementation of PA-practices relies on the teacher's ability to adequately prepare students for this activity. However, current research on PA offers little input on what the role of the teacher within these practices ought to be. Hovardas, Tsivitanidou and Zacharia (2014) point out that guidance and support throughout the whole process might be crucial. The tension for the teacher in this regard lies in finding the balance between on the one hand leaving the expert role as a teacher-assessor and on the other hand passing the process completely into the hands of the students. Previous research indicates that students used to a teacher-led assessment can become frustrated (Gielen, Dochy, Onghena, Janssens, & Decuyper, 2007) or experience discomfort (Vanderhoven et al., 2015) when the teacher assessment and feedback

opportunities are completely replaced by peer assessment and feedback. In a recent study by Güler (2016) students' preference for additional teacher assessment was confirmed. However, this recent study did not investigate students' perceptions when teacher's assessment and feedback is given anonymous and thus intertwined with the assessment and feedback of peers, compared to a setting in which the teacher's assessment and feedback is given non-anonymously. As MRT offers the possibility to create both aforementioned settings and little is known about students' perceptions on this issue, both the anonymous and non-anonymous setting for teacher assessment and feedback will be investigated in this study.

Research questions and hypotheses

As PA can be operationalized in many forms (Gielen, Dochy, & Onghena, 2011) it is important to clearly describe the adopted PA- mode when reporting on PA-research. In this study we focus on *reciprocal* (students assess each other and all students undertake both the role of assessor and assessee), *anonymous* (single blind: anonymity is offered to the assessor), *synchronous* (face-to-face) peer assessment of *group* presentations. A within-subjects quasi-experimental design study was conducted in higher education.

First, MRT will be investigated as a possible solution for facilitating both immediate peer assessment and feedback in class. A second aim was to investigate how the interpersonal processes are related to students' conceptions of PA (CON PA) and to students' perceived learning in PA (CONL PA). As mentioned in the theoretical framework, this part of the study builds on the results of the model of interrelatedness of interpersonal variables by van Gennip et al. (2010) and recent insights from the field. Based on the theoretical framework the hypotheses are depicted in Figure 1. Third, in search of the complementarity of peer feedback and teacher feedback in PA settings and taking into account the possible identity revelation modes that MRT offers, students' preferences towards (non-) anonymous teacher assessment and feedback will be studied. Following research questions are put forth:

RQ 1: How do students experience anonymous PA with the use of MRT?

RQ 2: What is the impact of interpersonal variables and conceptions of peer assessment in relation with perceived learning in PA?

RQ 3: What type of teacher assessment and feedback do students prefer within this PA setting?

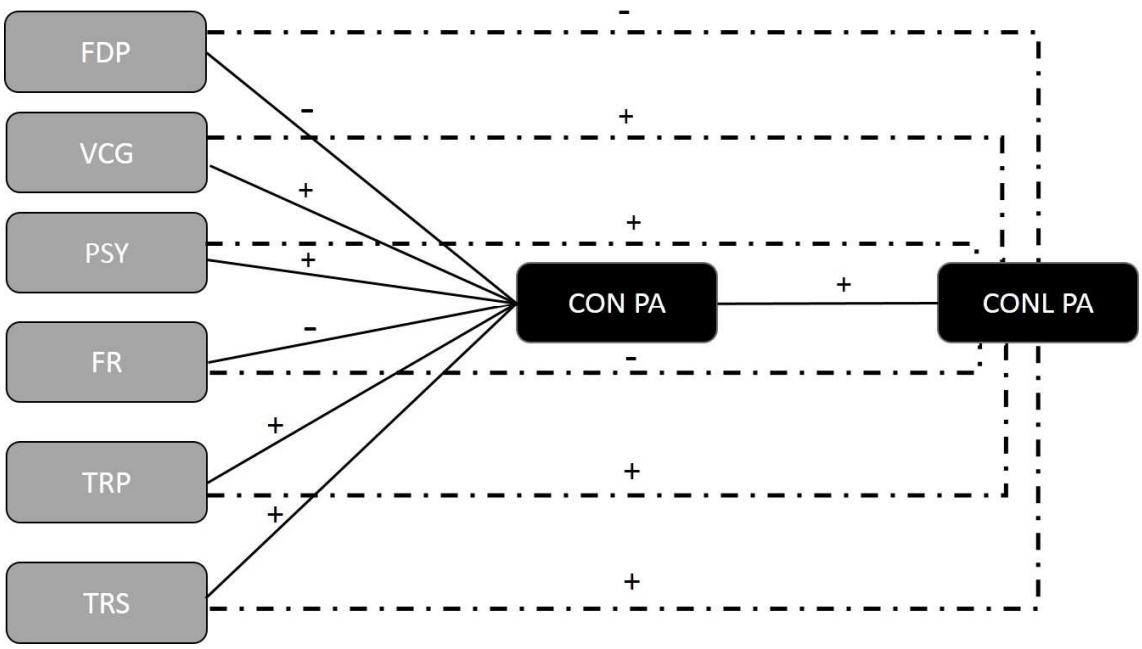


Figure 1. Conceptual Model

Legend: FDP = Fear of Disapproval; VCG = Value Congruency; PSY = Psychological Safety; FR = Friendship; TRP = Trust in Peer; TRS = Trust in Self; CON = General conceptions of PA; CONL PA = Conceptions of learning in PA

Method

Participants and setting

Participants in this study were 39 third-year bachelor students in Educational Studies who were enrolled in the course Instructional Design (Mage = 21.4). The majority were female (94.87%). The main objective of this course was to become acquainted with various instructional design strategies based on a social constructivist learning approach through authentic tasks (Tillema, Leenknecht, & Segers, 2011). Besides several theoretical lectures, students received a group assignment to prepare and present a workshop on one of the provided topics (e.g., The Jigsaw Classroom, Student Teams-Achievement Divisions, Teams-Games-Tournaments, Group Investigation and Peer Tutoring). Their workshop could last a maximum of 30 minutes. Furthermore, students were expected to develop learning materials in order to help their peers master these new topics. Teams consisted mostly of two or three students and were self-chosen (i.e. 4 teams of two students, 10 teams of three students and one team of 1 student due to a work status agreement). Students were asked to apply the pedagogical principles they had learned in the lectures (e.g.; principle of

gradualism) as much as possible. These principles were also used to peer-evaluate the workshops given by the various teams via MRT. Additionally, students were individually peer-evaluated on their presentation skills via three criteria (body language, use of voice and interaction with audience). The free available MRT-tool Socrative was used for this. The complete list of the evaluation criteria with 4 content-related criteria (i.e. pedagogical principles) and 3 presentation-related criteria (i.e. presentation skills) is shown in Appendix A.

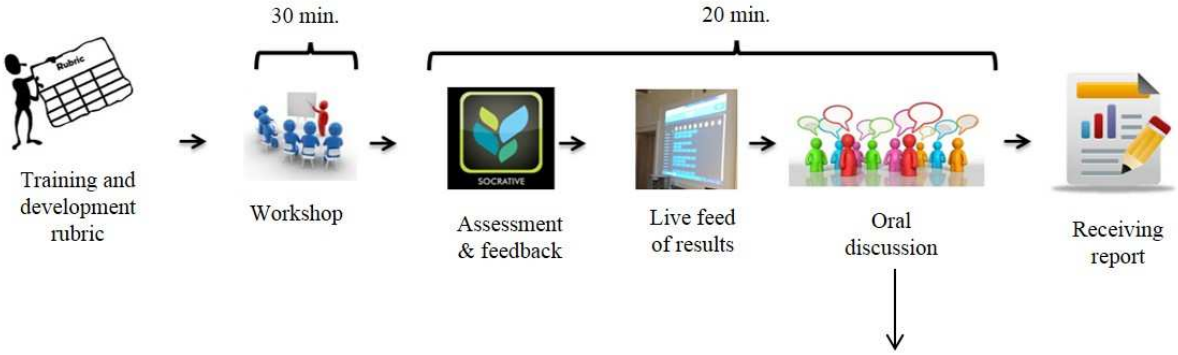
The function of the PA-activity was formative in nature as it was the university teacher's intention to let the students learn from their peers' input in light of future likewise performances. However, to stimulate effort and justify the investment of time that was put into development of the workshop 25% of the total grade for this course was awarded to the quality of the workshop. This grade was calculated as the mean PA-rubricscore for all criteria, including the rubricscores of the teacher. No marks were given for the quality of the peer feedback due to the specific focus of this study and the guaranteed anonymity for the assessors. At the beginning of the course students were asked by the university teacher whether they had previous experience with PA in higher or secondary education. Although all students had experience with PA by means of grading, none of them had experience with peer feedback, nor with MRT.

Procedure and design

As the quality of a PA-process (i.e. fairness and meaningfulness) can be substantially improved when comprehensive training is provided and when a rubric is used, students were given a multifaceted training (Liu & Li, 2013; Panadero et al., 2013; Planas Lladó et al., 2014; Ploegh, Tillema, & Segers, 2009).

The first part of the training consisted of a theoretical background comprising discussions of the advantages and disadvantages of PA and an introduction to the pedagogical principles. Second, in this introduction to the pedagogical principles, examples of expert presenting performance (i.e. modelling) were shown and assessed by the students via an example rubric that was used in a previous PA project (Fastré, van der Klink, Sluijsmans, & van Merriënboer, 2013). As a third part of the training, the students were divided in two similar groups, group A (n = 20) and B (n = 19), respectively consisting of 7 and 8 teams, and were assigned to formulate a rubric in accordance with one of the four pedagogical principles. Subsequently, students discussed each other's rubrics (separately in group A and B) and were given the opportunity to practice these rubrics via examples they brought themselves. Finally, the university teacher and assistant consolidated the

different rubrics of group A and B for the four pedagogical principles and three criteria related to individual presentation skills into one rubric that would be used in each workshop. The final rubric comprised descriptions of observable behaviour for every possible score on every criterion (scale from 1 to 5). An example is shown in Appendix B. Afterwards, during four consecutive weeks, students gave their workshops in parallel workshop sessions (group A and B). This means that all students attended and assessed half of the presented workshops (i.e. 8).



Conditions		Non-anonymous teacher assessment & feedback (NATAF)	Anonymous teacher assessment & feedback (ATAF)
Week	**Week 1 & 2**	Group B	Group A
	Week 3 & 4	Group A	Group B

Figure 2. Intervention outline and manipulation

The planned assessment and feedback episode after each workshop session included the following steps (see Figure 2): First, students and the teacher gave anonymous scores and feedback through Socrative on the given rubric scores via an electronic device in a WiFi-enabled classroom (laptop, tablet or own smartphone), and, the general results were projected live onto a Smartboard (live feed). Second, after all students had responded, the given results were orally discussed with the peers (oral discussion). The role of the teacher was to facilitate the discussion on the strengths and weaknesses of the workshops. This included both content related input to enforce shared understanding of the criteria (e.g., 'Why do you think the lack of qualitative sources is criticised?') as well as social-affective input, involving acts that build up trust and scale up mutual support between assessors and assessees (e.g., Do not worry too much when your timing was criticised, we all know it takes a lot of experience to get this right. Envision this task as a practice opportunity.') (Xu & Carless, 2016).

To be able to investigate students' preferences regarding the judgement of the teacher, a within-subjects design manipulation was set up, in which students experienced (from an assessor's perspective) two types of teacher assessment and feedback (i.e. *anonymous teacher assessment and feedback - ATAF* vs. *non-anonymous teacher assessment and feedback - NATAF*) during two consecutive sessions (see Fig. 2). In the *anonymous setting (ATAF)*, the teacher entered her scores and feedback anonymously via the MRT-tool and only moderated the oral peer feedback discussion. In the *non-anonymous setting (NATAF)*, the protocol was the same, with the exception that the teacher explicitly gave her scores and feedback during the oral peer feedback discussion on a criterion-by-criterion basis (i.e. she made the given scores and feedback she entered into Socrative explicit). In order to control for sequence effects, the order of experienced settings was different for both groups.

These two steps took about 20 minutes per workshop. Within a day, the Socrative reports (automatically generated Excel files) were sent to the assessed group. The report offers an overview of the anonymous assessment and feedback for both the content-related (on the group level) and presentation-related criteria (on the individual level).

It is important to note that when assessing other peers' work, students experienced both settings, but as assessees, they only experienced one type of teacher assessment and feedback (i.e. they only had to organise one workshop).

Measurements and analyses

Due to the specificity of the PA setting, all data for RQ 1 and 2 were gathered via post-questionnaires and complementary student reflection notes with open-ended questions. To answer RQ3 a pretest-posttest design was used to measure students' preference towards the teacher's judgement. A qualitative analysis of the answers to the open questions was carried out. This mixed-method approach allows us to obtain a complete picture of the intervention. All students were expected to take part in the PA-activities as developing feedback skills was mentioned as one of the learning goals of the course. A deductive thematic analysis started from 13 codes based on the scales of the quantitative data (e.g. *friendship marking, psychological safety, role of teacher, general conceptions to PA*). At the beginning of the course students filled out an informed consent. Independent t-tests were used to explore differences between group A and group B. Only when significant differences were found, these will be reported in the result section. In the following sections the measurements are discussed separately.

Research question 1: Experience with MRT. To measure whether the use of Socrative and its features were perceived as a valuable facilitator for PA and peer feedback practices, we need to know students' general conceptions of PA, their experience with the anonymous peer assessment and feedback setting, the perceived added value offered by the written feedback feature in Socrative, and their feelings about the fact that their scores and feedback are being live-projected in front of the whole class. The scales in the first part of the questionnaire were based on Raes et al. (2013) for the interpersonal variables and general conceptions of PA and on MacGeorge et al. (2008) for the technology-related questions. These are presented in Table 1, together with an example item of each scale and the corresponding Cronbach's α. All items were measured using a 7-point Likert scale from 1 (completely disagree) to 7 (completely agree). Furthermore, students were asked an open-ended question about their experiences with receiving scores and feedback via Socrative: *Do you think this tool is useful for collecting peer feedback?*

Table 1
Scales to answer Research Question 1

Students' perceptions regarding:	**Number of items**	**Cronbach's α**
General conceptions of PA *e.g. Peer assessment is useful*	8	.72
Importance towards anonymity *e.g. It is important that the sessions were anonymous.*	3	.72
Usability of Socrative *e.g. Socrative is accessible to use.*	5	.74
Added value of peer feedback feature *e.g. The possibility to provide arguments for a given score has an added value.*	3	.68
Feelings concerning appearance of scores & feedback (asked from an assessee's point of view) *e.g. The fact that everyone can see my scores and feedback on the screen, makes it transparent.*	4	.68

Research question 2: Relationship between interpersonal variables and conceptions of PA. As mentioned in the literature section, the scales for both the independent and dependent variables were based on Cheng and Tsai (2012), van Gennip et al. (2010) and Vanderhoven et al. (2015). Again a 7-point Likert scale was used. An example item of each scale is shown in Table 2. The following open-ended question was posed with regard to research question 2: *Do you think that the anonymous tool can diminish undesirable social effects such as distrust in each other's evaluative capabilities and friendship marking?*

Table 2
Scales regarding Research Question 2

Students' perceptions regarding…	**Number of items**	**Cronbach's α**
Fear of disapproval when giving low rubricscores/critical feedback (FDP) e.g. *When I give low rubricscores, I'm afraid that my peers would be offended.*	3	.89
Value congruency goals & criteria (VCG) e.g. *All students interpreted the goals of the PA and PF sessions in the same way.*	8	.73
Psychological safety (PS) e.g. *I feel comfortable giving my opinion in this group*.	4	.82
Friendship marking (FR) e.g. *I had the feeling that I had to give good scores and feedback to my friends.*	5	.71
Trust in the self (TRS) *e.g. I am capable of giving my friends feedback.*	7	.92
Trust in peers (TRP) e.g. *My peers are capable of giving an objective opinion on my presentation via rubricscores.*	7	.95
Conceptions of perceived learning in PA (CONL PA) *e.g. PA is a way to evaluate different perspectives on performances*	5	.60

Research question 3: Preferred type of teacher assessment and feedback. To answer the third research question, students' perceptions of the role of the teacher within PA-practices were measured before and after the intervention with Socrative (see Table 3). These perceptions were measured via a 4-item scale. Due to the two modes that were created for this third research question, two different questions from an assessee's point of view were asked depending on whether their presentation was scheduled in the anonymous teacher assessment and feedback setting (ATAF) or the non-anonymous teacher assessment and feedback setting (NATAF). Subsequently, one open-ended question was posed regarding research question 3: *What should be the role of the teacher during these sessions?*

Table 3
Scale and items regarding Research Question 3

	Items	**Cronbach's α pre-test**	**Cronbach's α post-test**
Perceptions of importance of teacher judgement in PA *e.g. I think it's important that the judgment of the teacher is explicitly communicated to the students* Specific questions anonymous teacher assessment and feedback setting (ATAF): *I experienced the fact that our group did not receive explicit feedback from the teacher as a deficiency.* *It is unfair that our group did not receive explicit feedback from the teacher.* Specific questions non-anonymous teacher assessment and feedback setting (NATAF): *If our group had not received explicit feedback from the teacher, I would have experienced this as a shortcoming.* *If our group had not received explicit assessment and feedback from the teacher, I would have found that unfair.*	4	*.75*	*.67*

Legend: *ATAF = Receiving anonymous teacher assessment and feedback / NATAF = Receiving non-anonymous teacher assessment and feedback

Results

RQ 1: How do students experience anonymous PA with the use of MRT?

The mean score on the scale *importance of anonymity* is moderately high for group A (M = 4.07, SD = 1.25) and high for group B (M = 5.62, SD = 1.165), $t(37) = 2.33$, $p = .00$, Cohen's d = .75. The majority of the participants (n = 35, 89%) expressed positive opinions on the offered anonymity, while the remaining 4 students stated that anonymity was not needed for them as they easily express their opinion, also in groups in which they do not know each other very well from the start. Student 14 expressed this positive opinion as follows:

> *The fact that our feedback was anonymous and as such it was not known who gave the feedback helped to give a substantiated score and feedback. If this had not been the case, my scores would have been different.*
> (Student 14, NATAF-condition)

Additional reasons for the positive attitudes towards anonymity were its effect on students' participation (n = 11, 28%) or fairness (n = 10, 26%), as reflected in the quotes by student 11 and student 25:

> *Everyone dares to type something anonymous instead of clarifying this vocally afterwards.*
> (Student 11, ATAF-condition)

> *Choosing an anonymous setting was the right choice. I was giving more honest scores and you know that the feedback that you will receive yourself will be more 'realistic'.*
> (Student 25, NATAF-condition)

Furthermore, the usability of MRT was evaluated very positively by all participants (M_{A+B} = 5.84; SD = .15; $t(37) = .86$, $p = .11$). Accordingly, in the open-ended questions, the majority of the students (n = 37, 95%) expressed their thoughts in this manner:

> *The Socrative tool was very user-friendly!*
> (Student 28, NATAF-condition)

Although I was rather sceptical in the beginning, Socrative has proven to be an easy-to-use tool which raises the interactivity level of these sessions.
(Student 15, ATAF-condition)

The two students that were not enthusiastic about use of MRT mentioned two reasons for their stance, that is, the first student argued that one is dependent on the MRT tool to organise the PA session and the second student thought that the oral discussion phase made the first anonymous phase redundant. Looking at the specific features of the MRT-tool in more detail, students highly valued the feedback feature in which they can give arguments for their scores (M_{A+B} = 6.16; *SD* = .648; *t*(37) = .30, *p* = .55). Most of the students (*n* = 28, 72%) expressed their idea about this as follows:

I think adding a motivation on the rubric criteria gives added value to the feedback.
(Student 10, ATAF-condition)

Feedback is offered in a clear and structured manner. The biggest advantage is that you get immediate feedback.
*(*Student 22, NATAF-condition)

Moreover, students appreciate the transparency that is offered via live projection and generally do not mention any feelings of discomfort herein (M_{A+B} = 5.03; *SD* = 1.01; *t*(37) = .16, *p* = .13). In the qualitative data, only one student mentioned a differing view on this element:

A possible disadvantage of anonymous judgement might be the fact that the commentaries appear 'uncensored'. I mean: some reactions were rather harsh. I assume that the person that gave these reactions via Socrative would not dare to do this in 'real life'. On the other hand, this also has advantages, because you don't have to be afraid to be honest. But even then, the articulation via Socrative can be read differently than if you would say it tactfully out loud.
(Student 13, ATAF-condition)

RQ 2: What is the impact of interpersonal variables and conceptions of peer assessment in relation with perceived learning in PA?

Means, standard deviations and Pearson correlations of variables measured are presented in Table 4. The results shows that students had very positive *general conceptions of PA* (M_{A+B} = 6.49; *SD* = .59; *t*(37) = -.95, *p* = .20). Moreover, results show that students scored high on *conceptions of learning in PA* (*M* = 5.24; *SD* = .68; *t*(37) = -1.00, *p* = .32). In addition, during the peer assessment and feedback sessions, we created an environment where students felt moderately safe for interpersonal risk-taking (M_{A+B} = 4.79; *SD* = .88; *t*(37) = -1.29, *p* = .87), and had a high amount of trust in their own (M_{A+B} = 5.01; *SD* = .82; *t*(37) = -.17, *p* = .87) and others' evaluative capabilities (M_{A+B} = 5.21; *SD* = .81; *t*(37) = .59, *p* = .56). For all variables in Table 4 no significant differences were found between group A and B.

Table 4.
Means, standard deviations and correlations for the measures in the research model.

	M	*SD*	1	2	3	4	5	6	7	8
1.Fear of disapproval (FDP)	3.16	1.36	-							
2.Value congruency goals & criteria (VCG)	4.69	.67	.01	-						
3.Psychological safety (PSY)	4.79	.88	-.44**	.46**	-					
4.Friendship marking (FR)	2.99	1.06	.64**	.12	-.39*	-				
5.Trust self (TRS)	5.01	.82	-.34*	.51**	.43**	-.17	-			
6.Trust peer (TRP)	5.21	.81	-.26	.65**	.46**	-.17	.79**	-		
7.General conceptions of PA (CON PA)	6.49	.59	-.58**	.43**	.42**	-.43**	.40*	.55**	-	
8.Conceptions of learning in PA (CONL PA)	5.24	.68	-.29	.51**	.39*	-.14	.47**	.56**	.47**	-

*Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

The influence of *friendship bonds* on students was moderately low (M_{A+B} = 2.99; *SD* = 1.06; *t*(37) = 1.26, *p* = .22), and this low score on the friendship variable was also reflected in the qualitative data. Two thirds of the students (*n* = 26) mentioned in their reflection notes that the fact that the sessions were anonymous for the assessors lowered the occurrence of friendship marking. This view was reflected, for instance, in the following comments:

> *Because of the anonymity I dared to give honest scores to my friends.*
> (Student 5, ATAF-Condition)

> *The anonymity within the Socrative setting was the biggest trump card. I had the feeling that this augmented the objectivity of the scores and helped in breaking down the barriers of shyness and manipulation when giving individual feedback. Even then -of course- I still tried to give feedback in a constructive way […]*
> (Student 23, NATAF-Condition)

However, one-third of the respondents (*n* = 13) mentioned that it is impossible to completely avoid friendship marking in peer assessment and feedback settings, although they are all convinced that anonymity can strongly dissuade students from doing this:

> *I think friendship marking will always play a role in these processes, simply by the fact that your sympathy for one peer is always bigger than for another. I personally think that you are more interested in the workshops of your close friends. You could give friends who didn't perform very well better scores, also because you know how much effort they put into the preparation of the workshop.*
> (Student 9, ATAF-condition)

> *The anonymous PA-tool indeed gives the possibility to diminish the effects of friendship marking, but it will never be able to banish it completely. You will always have people who unconsciously give better scores to their friends.*
> (Student 14, ATAF-condition)

Regarding the variable *value congruency*, the majority of the students agreed upon the goals and criteria of PA-sessions (M_{A+B} = 4.69; *SD* = .67; *t*(37) = .44, *p* = .67). Scores on the variable *fear of disapproval* when giving low scores were rather low (M_{A+B} = 3.16; *SD* = 1.36; *t*(37) = 1.10, *p* =

.28). This finding was confirmed by the fact that 8 students (20.5%) explicitly mentioned this variable when commenting on the importance of anonymity:

> *[…] Because in this [anonymous] way you might be more honest and can give a more subtle opinion score and feedback, especially to friends, and you don't have to be afraid that they will be mad at you […]*
> (Student 6, ATAF-condition)

Regarding the relation between interpersonal variables and general conceptions to PA (CON PA) a significant positive correlation was found for psychological safety (PSY) (r = .42, p < .01), value congruency on goals and criteria (VCG) (r = .43, p < .01), trust in peers (TRP) (r = .55, p < .01) and trust in self (r = .40, p < .05). Furthermore, a significant negative correlation was found CON PA and fear of disapproval (FDP) (r = -.58, p < .05) and friendship marking (FR) (r = -.43, p < .01) respectively. Concerning the relation between conceptions of learning in PA (CONL PA) and interpersonal variables, a significant positive correlation between psychological safety (PSY) (r = .39, p < .05), value congruency on goals and criteria (VCG) (r = .51, p < .01), trust in peers (TRP) (r = .56, p < .01*)* and trust in self (r = .47, *p < .01)* was found. Linear regression analyses were used to investigate the relationship between the interpersonal variables and the two dependent variables (CON PA and CONL PA). As can be seen in Figure 3, first, all independent variables were entered into the model in order to predict general conceptions of PA (CON PA). Dotted lines represent non-significant results. Results indicate that in an anonymous face-to-face peer assessment and feedback setting, *value congruency* (VCG) of the goals of the PA-sessions (β = .32, *std. error* = .15, p < .05) and *trust in peers'* evaluative capabilities (TRP) (β = .31, *std. error* = .16, p < .05) are positive predictors of *general conceptions of PA.* On the other hand, *fear of disapproval* (FDP) (β = -.22, *std. error* = .07, p < .01) negatively predicted *general conceptions of PA* (CON PA) (adjusted R^2 = .52). The other interpersonal variables were not found to predict conceptions of PA.

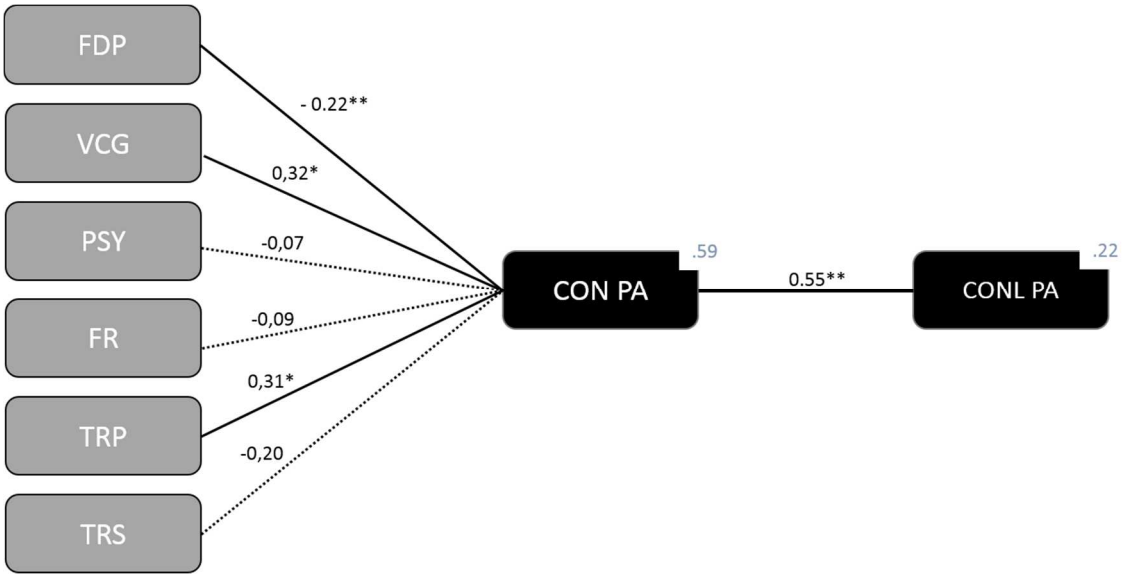


Figure 3. Tested model towards Conceptions of Learning in PA

Legend:

FDP = Fear of Disapproval; VCG = Value Congruency; PSY = Psychological Safety; FR = Friendship; TRP = Trust in Peer; TRS = Trust in Self; CON = General conceptions of PA; CONL PA = Conceptions of learning in PA
* p < .05 / ** p < .01

In a second analysis, all interpersonal variables were entered to predict *conceptions of learning in PA* (CONL PA). None of the interpersonal variables were significant predictors of CONL PA (adjusted R^2 = .22).

In order to explore a possible mediating effect of CON PA between the interpersonal variables and CONL PA, a path analysis was conducted using the R-software package Lavaan (Rosseel, 2012). Due to the large-model conditions and a rather small sample size, a Swain correction was calculated using Boomsma and Herzog's (2012) R-function, in order to obtain proper estimates of the model statistics and indexes. Results in Figure 3 show that CON PA is a highly positive predictor of CONL PA (β = .55, p < .01). This model explains 22% of the variance of the dependent variable CONL PA (Swain corrected χ^2 = 9.08, p > 0.05; Swain-corrected CFI= .89; Swain-corrected TLI= .47; Swain-corrected RMSEA = .12).

RQ 3: What type of teacher assessment and feedback do students prefer within this PA setting?

The results show that students highly prefer explicit teacher assessment and feedback when involved in anonymous PA (M_{Pre} = 5.53, SD = .88; M_{Post} = 5.63, SD =.85; t(38) = -.50, p = .62).

Moreover, no significant difference was found between the students who received non-anonymous teacher assessment and feedback and those who did not (M_{NATAF} = 5.38, *SD* = .95; M_{ATAF} = 5.86, *SD* = .656; *F*(1,38) = 3.43, p = .07). Twelve respondents (31%) refer directly to the added value of the teacher's expertise in evaluating when expressing their preferences regarding the teacher role within these practices:

> *Due to the objectivity and the expertise of the teacher, I consider her feedback and score to be important. On the other hand, it is good that this is not the only source of feedback and assessment, but I do think it should be present.*
> (Student 17, ATAF-Condition)

This finding is confirmed by the responses to the questions posed from an assessee's perspective. The teams that did not receive explicit teacher assessment and feedback experienced this as a great deficiency (*M*= 6.55; *SD* = .61). Additionally, the groups that did receive explicit teacher assessment and feedback stated that not receiving teacher assessment and feedback would be a critical shortcoming in the assessment and feedback procedure (*M* = 5.56; *SD* = 1.34). Furthermore, several students (*n* = 10, 26%) also mentioned that the assessment and feedback from the teacher was used to crosscheck with the assessment and feedback from peers. Students used this input to determine how they valued the assessment and feedback from peers:

> *[…] Also getting input from the teachers makes it possible to compare her opinion with the input from peers, in this manner you can decide what peer feedback you're willing to take along.*
> (Student 17, ATAF-Condition)

> *[…] The feedback from peers is sometimes contradictory, and when you receive a score from the teacher, this determines how you experience criticism!*
> (Student 38, NATAF-Condition)

Interestingly, when commenting on what the role of the teacher should be in this PA-setting, some students (*n* = 7, 18%) also mentioned that over time, during the oral discussion phase that was moderated by the teacher, the group atmosphere became 'more open' and students participated more actively in this phase.

> *[…] the oral discussions were sometimes confrontational, but they are indispensable. In the first session I felt unpleasant when giving oral feedback to peers, but as time passed and due to the open atmosphere in the group, this became easier to do.*
> (Student 35, NATAF-Condition)

Discussion and conclusion

This study investigated the use of Mobile Response Technology in in-class peer assessment and feedback sessions in higher education (research question 1). Moreover, the study studied the relation between interpersonal variables and general conceptions of PA and conceptions of perceived learning in PA in an anonymous setting, in order to gain insight into their predictive value (research question 2). Finally, students' preferred type of assessment and feedback by the teacher within this PA-setting was examined (research question 3).

First of all, we can conclude that students in this study had positive conceptions of peer assessment and feedback in general, which indicates that the introduction to the process of peer assessment and the involvement in the development of the rubric was successful (Planas Lladó et al., 2014). Regarding research question 1, the MRT-tool is appreciated for its user-friendly applicability and its immediate written feedback feature, which is seen as a key element in PA-practices (Nicol, 2010). More specifically, students in this study value the transparency that is offered via the live projection and, from an assessee's point of view, do not mention any feelings of discomfort herein. The fact that MRT is designed for educational purposes makes it an easier tool to use for peer assessment and peer feedback, compared to the recent exploration of the use of WhatsApp for these purposes by Güler (2016). Features as the projection of summaries of students' scoring input via histograms, the well-structured display of the peer feedback and the ability to generate feedback reports, clearly decrease the in-class time as well as the effort for educators to get acquainted with the tool. Moreover, as also stressed by Vanderhoven et al. (2015), students value the high level of guaranteed anonymity for the assessors, and the qualitative data show that this adds to the (face) validity of the assessment process.

With regard to research question 2, the results show that a PA-setting was created in which students felt safe in terms of interpersonal risk-taking, and had a high amount of trust in their own and others' evaluative capabilities. Moreover, there was a high unanimity on the goals and the criteria of the PA-task. The majority (66%) explicitly mentioned that thanks to the guaranteed anonymity, they gave more honest scores and feedback and were less influenced by friendship bonds. This was also reflected in the quantitative results. However, one-third of the respondents believe that anonymity only offers a diminishing effect, and that friendship marking will always

be present within PA. A more "fine-grained operationalization of friendship" (Panadero et al., p.201, 2013) would offer a greater understanding of the issue of friendship marking.

The scores on the variable *fear of disapproval when giving low scores or critical feedback* were rather low, and one fifth of the participants directly relate their low score on this variable to the guaranteed anonymity for the assessor, thus diminishing fear for repercussions. This confirms similar findings by Vanderhoven et al. (2015). Regarding the relations between all interpersonal variables and general conceptions of PA within an anonymous PA-setting, firstly the results show that when students report a high amount of fear of disapproval, they show more negative general conceptions of PA. Secondly, the higher the unanimity of goals and criteria of PA, the more positive conceptions of PA students have. This finding is in contrast with the findings of van Gennip et al. (2010), who found that the lower the value congruency on goals, the more positively students think about PA in general. The anomaly in their findings was therefore confuted in the current study. Likewise, a higher amount of trust in the evaluative capabilities of peers predicts more positive conceptions of PA. Furthermore, students' general conceptions of PA seem to be a strong indicator of their conceptions of learning in PA. The mediating role of general conceptions of PA in the study by van Gennip et al. (2010) was confirmed.

The results of the third research question revealed no differences between students' preferences regarding teacher's explicit communication about their scores and feedback before and after the PA-sessions. Likewise, no differences could be found between the two teacher assessment and feedback conditions in this study: in both the anonymous and non-anonymous teacher assessment and feedback condition, participants wished to receive clearly identifiable and explicit teacher assessment and feedback. Teacher's expertise and perceived higher objectivity were mentioned as the main reasons for this preference. Similarly to recent findings by Hovardas et al. (2014) and Tsivitanidou et al. (2016), the qualitative data indicate that students tend to use a cross-checking technique between the peer and the expert feedback in order to deal with contradictory peer assessment and feedback, and filter which feedback they take with them for future performances. It seems that both the teacher's moderation of the oral discussion phase as well as the teacher's feedback serves multiple goals (Xu and Carless, 2016). It is up to future research to deepen our understanding on the feedback enabling processes that teachers can initiate during this oral discussion phase, as recently explored by Xu et al. (2016), that is both providing high quality feedback and preparing student to be social-affectively ready for feedback processes, both in the assessee and assessor role.

Another important finding is that in a PA-setting in which the possibility for immediate anonymous written feedback is offered, students still prefer to have the possibility for non-anonymous oral feedback in the PA procedure. In this phase peers can voluntarily comment on

the given feedback via MRT, and the assessees have the opportunity to respond. However, it took several sessions for students to actively participate in the oral discussion part. In the qualitative data, some students mentioned that the group atmosphere became more open as the PA-sessions went on. It might be that the combination of an anonymous and non-anonymous phase within peer assessment and feedback settings is the key to create a learning environment where students have the opportunity to exchange their interpretations and are able to negotiate meaning in safe learning contexts, as advocated by Nicol (2010) and called 'dialogic feedback' by Carless et al. (2011). Future research should gain insight into the time factor within these processes, preferably within different educational contexts, for example through fading anonymity over time, and in this manner gradually evolving towards authentic non-anonymous peer feedback situations.

It might be argued that we are only facing a novelty effect, namely, students like MRT since it is new technology used in the classroom. In this study, this effect would disappear if MRT would be regularly used. However, Draper and Brown (2004) found that this novelty effect exists for the use of clickers in classes, but only lasts between 5 and 50 min. After this period there is more chance for an anti-novel resulting in students being more unconvinced. Since our participants have used the MRT-tool for several sessions spread over 4 weeks, there is no reason to believe our positive effects would fade away with more time. Overall, based on the mean rubric scores, all students performed reasonably well, which might have influenced the students' opinion on the whole PA experience. Furthermore, all participants were students in educational studies who were familiar with mobile technologies. The aforementioned limitations urge for conducting similar studies with the participation of students from different departments and other educational levels.

Due to the guaranteed anonymity it was impossible to compare students' questionnaire responses to their actual scoring and feedback behaviour. Future research on anonymity within PA-settings should design quasi-experimental settings which enable such a comparison. As a consequence, no content analysis of the peer feedback could be performed.

In conclusion, this study showed that Mobile Response Technology can be an efficient and effective facilitator when organising a face-to-face peer assessment and feedback practice within higher education. Indications were found that providing anonymity for the assessors might result in a reduction of possible negative social effects generated by interpersonal variables. Accordingly, students are offered a safe environment in which to practise evaluative judgement in order to develop their assessment skills (Boud, Lawson, & Thompson, 2013). When peer assessment and feedback is given anonymously students prefer explicit teacher assessment and feedback, as it can be juxtaposed with contradictory feedback given by peers.

References

Ainsworth, S., Gelmini-Hornsby, G., Threapleton, K., Crook, C., O'Malley, C., & Buda, M. (2011). Anonymity in classroom voting and debating. *Learning and Instruction*, *21*, 365–378. doi:10.1016/j.learninstruc.2010.05.001

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing Procedures for Implementing Peer Assessment in Large Classes Using an Action Research Process. *Assessment & Evaluation in Higher Education*, *27*(5), 427–441. doi:10.1080/0260293022000009302

Boomsma, A., & Herzog, W. (n.d.). Correcting structural equation model fit statistics and indexes under small-sample and/or large-model conditions. Retrieved from http://www.gmw.rug.nl/~boomsma/swain.pdf

Boud, D., Lawson, R., & Thompson, D. G. (2013). Assessment & Evaluation in Higher Education Does student engagement in self- assessment calibrate their judgement over time ? *Assessment & Evaluation in Higher Education*, 37–41. doi:10.1080/02602938.2013.769198

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE-Life Sciences Education*, *6*, 9–20.

Carless, D. (2013). Trust and its role in facilitating dialogic feedback. In *Feedback in Higher and Professional Education* (pp. 90–103). doi:10.4324/9780203074336

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*. doi:10.1080/03075071003642449

Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education*, *35*(5), 551–564. doi:10.1080/02602931003632381

Cheng, K. H., & Tsai, C. C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology*, *28*, 599–618.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning*, *20*, 81–94. doi:10.1111/j.1365-2729.2004.00074.x

Fastré, G. M. J., van der Klink, M. R., Sluijsmans, D., & van Merriënboer, J. J. G. (2013). Towards an integrated model for developing sustainable assessment skills. *Assessment & Evaluation in Higher Education*, *38*(5), 611–630. doi:10.1080/02602938.2012.674484

Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education*, *36*, 137–155. doi:10.1080/02602930903221444

Gielen, S., Dochy, F., Onghena, P., Janssens, S., & Decuyper, S. (2007). A complementary role for peer feedback and staff feedback in powerful learning environments. In S. Gielen (Ed.), (pp. 157–199). Leuven: Katholieke universiteit Leuven,Faculteit Psychologie en Pedagogische Wetenschappen Centrum voor Opleidingsdidactiek.

Güler, C. (2016). Use of WhatsApp in Higher Education: What's Up With Assessing Peers Anonymously? *Journal of Educational Computing Research*. doi:10.1177/0735633116667359

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, *36*, 101–111. doi:10.1016/j.tate.2013.07.008

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, *38*, 21–27. doi:10.1016/j.stueduc.2012.04.001

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers and Education*, *71*, 133–152. doi:10.1016/j.compedu.2013.09.019

Howard, C. D., Barrett, A. F., & Frick, T. W. (2010). Anonymity to Promote Peer Feedback: Pre-Service Teachers' Comments in Asynchronous Computer-Mediated Communication. *Journal of Educational Computing Research*, *43*(1), 89–112. doi:10.2190/EC.43.1.f

Hunsu, N. J., Adesope, O., & Bayly, D. J. (2016). A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Computers and Education*, *94*, 102–119. doi:10.1016/j.compedu.2015.11.013

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, *20*, 344–348. doi:10.1016/j.learninstruc.2009.08.005

Lin, S. s. j., Liu, E. z. f., & Yuan, S. m. (2001). Web-based peer assessment: feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, *17*(4), 420–432. doi:10.1046/j.0266-4909.2001.00198.x

Liu, X., & Li, L. (2013). Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment. *Assessment & Evaluation in Higher Education*, *0*(0), 1–18. doi:10.1080/02602938.2013.823540

Lu, R., & Bol, L. (2007). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning*, *6*(2), 100–115.

MacGeorge, E. L., Homan, S. R., Dunning Jr, J. B., Elmore, D., Bodie, G. D., Evans, E., … Geddes, B. (2008). Student evaluation of audience response technology in large lecture classes. *Educational Technology Research and Development*, *56*, 125–145.

Magaña, S., & Marzano, R. J. (2014). Using Polling Technologies to Close Feedback Gaps. *Educational Leadership*, 82–83.

Nicol, D. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, *35*(5), 501–517. doi:10.1080/02602931003786559

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Human factors and social conditions of assessment. New York: Routledge* (pp. 1–39). New York, NY: Routledge.

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, *39*(4), 195–203. doi:10.1016/j.stueduc.2013.10.005

Planas Lladó, A., Soley, L. F., Fraguell Sansbelló, R. M., Pujolras, G. A., Planella, J. P., Roura-Pascual, N., … Moreno, L. M. (2014). Student perceptions of peer assessment: an interdisciplinary study. *Assessment & Evaluation in Higher Education*, *39*(5), 592–610. doi:10.1080/02602938.2013.860077

Ploegh, K., Tillema, H. H., & Segers, M. S. R. (2009). In search of quality criteria in peer assessment practices. *Studies in Educational Evaluation*, *35*(2–3), 102–109. doi:10.1016/j.stueduc.2009.05.001

Raes, A., Vanderhoven, E., & Schellens, T. (2013). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Studies in Higher Education*, *0*(0), 1–16. doi:10.1080/03075079.2013.823930

Rosseel, Y. (2012). Lavaan: An R package for structural equation modelling. *Journal of Statistical Software*, *48*(2), 1–36.

Sluijsmans, D. M. A., & Brand, S. (2002). Assessment & Evaluation in Higher Education Peer Assessment Training in Teacher Education : Effects on performance and perceptions. *Assessment & Evaluation in Higher Education*, *27*(5), 443–454. doi:10.1080/026029302200000931

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Martens, R. L. (2004). Training teachers in peer-assessment skills: effects on performance and perceptions. *Innovations in Education and Teaching International*, *41*, 59–78. doi:10.1080/1470329032000172720

Stowell, J. R. (2015). Use of Clickers vs. Mobile Devices for Classroom Polling. *Computers & Education*. doi:10.1016/j.compedu.2014.12.008

Strijbos, J. W., Ochoa, T. A., Sluijsmans, D. M. A., Segers, M. S. R., Tillema, H. H. (2009). Fostering interactivity through formative peer assessment in (web-based) collaborative learning environments. In *In C. Mourlas, N. Tsianos, & P. Germanakos (Eds.), Cognitive and emotional processes in web-based education: Integrating human factors and personalization.* (pp. 375–395). Hershey, PA: IGI Global.

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, *20*, 265–269.

Sung, Y.-T., Chang, K.-E., & Liu, T.-C. (2015). The Effects of Integrating Mobile Devices with Teaching and Learning on Students' Learning Performance: A Meta-Analysis and Research Synthesis. *Computers & Education*, *94*, 252–275. doi:10.1016/j.compedu.2015.11.008

Tillema, H., Leenknecht, M., & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning – A review of research studies. *Studies in Educational Evaluation*, *37*(1), 25–34. doi.org/10.1016/j.stueduc.2011.03.004

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*(4), 339–343. doi:10.1016/j.learninstruc.2009.08.003

Tsivitanidou, O. E., & Constantinou, C. P. (2016). A study of students' heuristics and strategy patterns in web-based reciprocal peer assessment for science learning. *The Internet and Higher Education*, *29*, 12–22. doi:10.1016/j.iheduc.2015.11.002

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, *4*, 41–54. doi:10.1016/j.edurev.2008.11.002

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, *20*(4), 280–290. doi:10.1016/j.learninstruc.2009.08.010

Van Zundert, M. J., Sluijsmans, D. M. A., Konings, K. D., & Van Merrienboer, J. J. G. (2012). The differential effects of task complexity on domain-specific and peer assessment skills. *Educational Psychology*, *32*, 127–145. doi:10.1080/01443410.2011.626122

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, *81*, 123–132. doi:10.1016/j.compedu.2014.10.001

Vickerman, P. (2009). Student perspectives on formative peer assessment: an attempt to deepen learning? *Assessment & Evaluation in Higher Education*, *34*, 221–230. doi:10.1080/02602930801955986

White, P., Syncox, D., & Alters, B. (2011). Clicking for grades? Really? Investigating the use of clickers for awarding grade-points in post-secondary education. *Interactive Learning Environments*, *19*(5), 551–561. doi:10.1080/10494821003612638

Xu, Y., & Carless, D. (2016). "Only true friends could be cruelly honest": cognitive scaffolding and social-affective support in teacher feedback literacy. *Assessment & Evaluation in Higher Education*, 1–13. doi:10.1080/02602938.2016.1226759

Yang, Y. F., & Tsai, C. C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction*, *20*, 72–83. doi:10.1016/j.learninstruc.2009.01.003

Yu, F.-Y., & Sung, S. (2015). A mixed methods approach to the assessor's targeting behavior during online peer assessment: effects of anonymity and underlying reasons. *Interactive Learning Environments*, 1–18. doi:10.1080/10494820.2015.1041405

Appendix A

Evaluation criteria used during the workshops

Content-related criteria:
1. Principle of illustration and specificity
2. Principle of activity
3. Principle of confinement
4. Principle of gradualism
Presentation-related criteria
5. Body language
6. Use of voice and Language
7. Interaction with the audience

Appendix B

Example A: Principle of gradualism

Score 1	Score 2	Score 3	Score 4	Score 5
There is a total lack of structure, which makes it impossible for the audience to understand the content. The group lacked time management.	There was an attempt to structure the workshop, but they failed. Time management was poor.	The material was logically built up, though its presentation lacked structure. The session ended on time, though the use of time wasn't efficient.	There was a good build-up and structure. The group presented the material in an effective way. Time management was good.	The workshop had a clearly structured build-up. Time management was optimal.

Example B: Interaction with the audience

Score 1	Score 2	Score 3	Score 4	Score 5
No (eye) contact with audience. No use of humour, very static. Lack of confidence.	Answers questions from the audience without going into them deeply. Little eye contact Corny humour, quite static,	Sometimes the audience is asked for an opinion, though it's not processed Occasionally some humour, adequately dynamic Sufficient eye contact	Gets the audience involved through concrete questions Adequate sense of humour Dynamic Constant eye contact (different persons)	Starts from the interest of the audience Discusses with the audience, makes them enthusiastic Highly confident speaker

5 ‘Now you know what you’re doing right and wrong!’ PF Quality in synchronous PA in secondary education

This chapter is based on:

Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2017). ‘Now you know what you’re doing right and wrong!’ Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education*. doi: 10.1007/s10212-017-0329-x

Chapter 5
'Now you know what you're doing right and wrong!' PF Quality in Synchronous PA in secondary education

Abstract

This study explores the effects of peer assessment (PA) practice on peer feedback (PF) quality of 11th grade secondary education students (N=36). The PA setting was synchronous: anonymous assessors gave immediate PF using Mobile Response Technology during 10 feedback occasions. The design was quasi-experimental (experimental vs. control condition) in in which students in one condition received a scaffold to filter out relevant information they received. It was expected that this filter-out scaffold would influence PF quality in subsequent tasks in which they were assessors. PF content analysis showed that offering multiple PF occasions improved PF quality: messages contained more negative verifications, and informative and suggestive elaborations after the intervention. However, no effects were found of filtering out relevant information on PF quality. Moreover, students' perceived peer feedback skills improved which was in correspondence with their actual quality improvement over time. Additionally, the perceived usefulness of the received feedback was rated high by all participants.

Introduction

Peer assessment (PA) is a peer-assisted, collaborative learning arrangement that includes students assessing their fellow students' performance by providing feedback, which could be quantitative in nature (i.e. grades or ratings across assessment criteria) and/or qualitative (written or oral comments) (Topping, 1998). The process of assessing and commenting on the strengths and weaknesses of peers' work can help familiarize the assessor with the evaluation criteria, and in this way help to develop knowledge on what constitutes good work, and what needs to be avoided (Yu & Sung, 2015). Numerous studies on PA have shown various benefits for the learning process, such as improved student motivation, improved conceptual understanding, communication skills and self-assessment skills (e.g. Falchikov & Goldfinch, 2000; Topping, 2009).

Peer Feedback (PF) is an important component of PA being the information that one student provides to a peer (e.g. Topping, 1998). Research emphasizes that students require practice and training to become skilled peer assessors and assessees, who provide and receive high-quality PF (Sluijsmans, 2002). Researchers and teachers are thus challenged to implement assessment

activities in which students are prompted to provide frequent PF, resulting in a frequent enactment of the peer-assessor role (Tsivitanidou & Constantinou, 2016). Additionally, it has been argued that more research is needed to explore the quality of PF in PA and how it is perceived by students, because PF does not automatically lead to positive outcomes (Shute, 2008). Furthermore, more insight is needed into the kind of support students need in order to improve the quality of the feedback they produce (e.g. Tsivitanidou, Zacharia, & Hovardas, 2011). Two different types of scaffolds for PA have been explored: (a) guiding questions (e.g. helping the assessee to "filter out" the relevant information), and (b) establishing criteria and use of rubrics. The aim of this study is twofold: exploring the effect of PA practice and PA scaffolds - especially "filter out"- on PF quality.

Training and practice as a prerequisite for a valuable peer assessment activity

The success of PA depends, to a great extent, on whether students are able to acquire critical assessment skills and provide valid judgement of their peers' work (Liu & Li, 2013) . Sluijsmans, Brand-Gruwel, van Merriënboer and Martens (2004) indicate that the general PA skill consists of three mains skills: (1) defining assessment criteria: thinking about what is required and referring to the products; (2) judging the performance of a peer: reflecting upon and identifying strengths and weaknesses; and (3) providing feedback for future learning. Previous research indicated that the development of the first and third assessment skill can be successfully trained (Sluijsmans, 2002; Van Zundert, Sluijsmans, & Van Merriënboer, 2010) and/or supported through assessment scaffolds (see further). However, it can be expected that the second assessment skill of being able to judge the performance of a peer requires multiple enactments in both the assessor-role and assessee-role. The more practice in these PA processes, the more likely students will develop the expertise for making sound PA judgements (Liu & Carless, 2006; Panadero, 2016). Additionally, PA practice enhances students' ability to produce higher work quality themselves (Sadler, 2010) as they will be better able to apply the internalised assessment criteria themselves. The importance of multiple enactment in PA is also acknowledged in Boud's view on sustainable assessment in which the capacity for students to make judgements of their own work is seen as essential to stimulate lifelong learning (Boud & Soler, 2015). Within this framework, PA offers students the opportunity to practice their evaluative judgements, which has simultaneously an impact on the peer assessors' self-regulating skills (Nicol, 2010).

In sum, previous research shows that practice and training are crucial for the development of PA skills (Sluijsmans, 2002). However, literature on how practice enhances PA skills is scarce and, furthermore, it has not been explored what are its effects on PF quality.

When is peer feedback of good quality?

The quality of PF is crucial as it is the basis for PA and provides a platform for engaging students in an interactive and elaborative feedback discourse as well as in taking ownership of their learning (Hattie & Gan, 2011). Additionally, PF has the advantage of bringing students in a situation in which they 'are on the same wavelength', making PF more understandable and useful for them (Topping, 2003). When writing feedback, students have more opportunities to engage in important cognitive activities such as critical thinking (i.e. deciding what contributes a good piece of work), planning, monitoring, and regulation (Lin, Liu, & Yuan, 2001). In essence, well-formulated feedback should provide an answer to three questions: 'Where am I going?' (feed up), 'How am I going?' (feedback) and 'Where to next?' (feed forward) (Hattie & Timperley, 2007). As stated in the recent work by Reinholz (2015) there are three broad categories of feedback: (1) process-focused, (2) product-focused and (3) person-focused feedback. Process-focused feedback encompasses both task feedback (i.e. whether or not the task is correctly completed) and self-regulation (i.e. how students monitor, self-control and direct their work during the task). Product-focused feedback relates to the (in)correctness of the task and why this is the case. Person-focused feedback is related to the person who is engaged in the task. Building on the recent work by Gielen and De Wever (2015) in this study we focus on product-focused feedback. We want to explore how students are able to improve their PF skills through practice and support of guiding questions in a synchronous anonymous PA setting in which immediate feedback is given (see further). Previous research indicated that qualitative feedback should provide two types of information: verifications and elaborations (Narciss, 2008). Verification refers to 'a dichotomous judgment to indicate that a response is right or wrong' and elaboration refers to 'relevant information to help the learner in error correction' (Hattie & Gan, 2011, p. 253). These types of information are thus seen as the structural components of feedback, because students require feedback that tells them not only if they performed the task correctly, but also why and what they should do about it to improve their work (e.g. Prins, Sluijsmans, Kirschner, & Strijbos, 2010). Therefore, offering elaborations that justify the verification (e.g. correct vs. incorrect) is presumed to be beneficial for students' learning. As a consequence a balanced proportion of verifications and elaborations is more valuable than just providing verifications alone (Gielen & De Wever, 2015).

Regarding the specific case of PF quality in PA settings, this has been explored in a series of recent studies by Gielen and De Wever (2012, 2015). These authors explored whether the use of guiding questions (e.g. "What would you change in your peer's work?") influenced PF quality. Regarding verifications, it was found that student usually tend to give mostly positive verifications (i.e. this is correct) in PA. However, the guiding questions used in the experimental conditions resulted in more negative verifications. According to the authors, this resulted into better PF

because it provoked a more balanced proportion of positive and negative verifications, and therefore it was more descriptive of the actual performance rather than just pointing out to the positive aspects as PA assessors tend to do. Regarding elaboration Gielen and De Wever found that the guiding questions effect resulted into more suggestive elaborations (i.e. feedback on how to improve a future performance) but did not have an effect on informative elaborations (i.e. feedback on why a criterion was performed correctly or not). Finally, no differences were noted between the proportion of informative and suggestive elaborations.

Importantly, Gielen and De Wever explored the effects of the guiding questions in PF quality with higher education students, practicing 3 times PA and in an asynchronous (i.e. non-immediate PF) wiki environment. In the present study we wanted to check whether the effect of guiding questions in improving PF quality would remain with secondary education students, when there is a stronger PA practice (10 PA occasions over an school year) and organized in a synchronous PA setting (i.e. immediate PF). Originally, Gielen and De Wever used a setting where the assessors were non-anonymous. However, in the present study anonymity for the assessors is assured. The rationale behind making the assessors anonymous is to decrease negative effects as a consequence of interpersonal processes (Panadero, 2016; Vanderhoven, Raes, Montrieux, Rotsaert, & Schellens, 2015).

How to support students to provide high quality peer feedback in peer assessment settings?

PA is often described as a complex collaborative learning task that requires high-level cognitive processing (e.g. Kollar & Fischer, 2010). Therefore, any approach to help students to provide better PF to their classmates will have an impact on the PA implementation and, finally, on learning. Previous research has explored two initiatives to support students in providing high quality PF. First, offering guiding questions/guidelines on what good PF quality constitutes (Reinholz, 2015). The logic behind is that, by offering such questions, the students will reflect more about the PA exercise which thus becomes a more metacognitive activity. This type of questions can be used to help PA assessors in producing the feedback and/or by assessees to better understand the feedback received. There are multiple examples of these types of interventions such as the previously presented work by Gielen and De Wever (2012, 2015), who provided a template for the assessors. Another suggested approach has been to help assessees to filter out the feedback they received (Tsivitanidou & Constantinou, 2016). The hypothesis behind this is that when assessees actively process the PF they receive, they will become better assessors in a subsequent task, which means that they will produce better quality PF. This hypothesis will be tested in the present study.

The second PA scaffold initiative involves the students in the selection of the PA assessment criteria and through the use of a rubric (Panadero et al., 2013; Sluijsmans, 2002). A rubric articulates the expectations for an assignment by listing the assessment criteria and by describing levels of quality in relation to each criterion (Reddy & Andrade, 2010). By using rubrics, students have a clearer understanding of what is expected of them as assessors and assessees, because rubrics provide assessment transparency (Panadero & Jonsson, 2013). A rubric is therefore often categorized as an assessment scaffold in PA research, one that has shown to increase the accuracy of PA (Panadero, Romero, & Strijbos, 2013). Therefore, rubrics are a positive support for PA. For that reason, all the participants in the present study will receive rubrics to enhance the potential of the PA tasks.

In sum, this study explores the effect of the first scaffolding approach, that is, the use of guiding questions by helping assessees to filter out the PA feedback. Two conditions will be compared: an experimental condition where students are actively supported to filter out PF vs. a control condition with no filter out scaffold. Additionally, both conditions help the assessors by providing them with guiding questions on how to assess and rubrics.

Importance of peer feedback skills and perceived usefulness

Identifying oneself as an active learner is a key element in the development of PF skills. For that reason, it is important to also incorporate students' perceived improvement of their feedback skills (Boud & Soler, 2015): if students perceive that they are becoming more capable as peer assessors, they will be more motivated to perform PA and believe it is useful. However, this has not been explored in detail in previous research. Therefore, we will provide 10 PA occasions so that the participants will have plenty of experience with PA and so that we can explore the evolution of their perceptions along with their veridicality (because it is checked whether they were actually becoming a better PA assessor).

Furthermore, the willingness to follow the assessors' advice is essential to augment the quality of the performance (Boud, 2000; Nelson & Schunn, 2009). How students respond to PF is not just a feature of the activities themselves; this depends also on the ways in which PF is perceived useful (i.e. mindful reception of PF) which cannot be controlled in advance (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991) . Therefore, students' perceived usefulness of the received PF will also be explored in this study.

Research questions & hypotheses

The aim of this paper is twofold. First, to analyze the effect of PA practice on PF quality over time. And, second, to explore the effect of a scaffold which helps assessees to filter out the feedback they received into providing PF of better quality in subsequent tasks. Additionally, it was explored whether students' perceived an improvement of their PF skills, which will be compared to their actual evolution. Finally, we explored whether the PF is perceived as useful from an assessee's point of view.

Given these research aims, a PA setting in which participants could take the roles of both assessor and assessee was needed. Hence we created a reciprocal PA setting where groups of students assessed each others' work. Additionally, this study is organized around two performance and multiple PA cycles to explore the effects of practice (10 PA occasions). Consequently, this allows us to measure the evolution of PF quality over time. The specific research questions and hypotheses are:

RQ 1: What is the evolution in PF quality over time when students practice PA several times?

(H1) It is expected to find a practice effect with an increase of negative verifications and suggestive elaborations (Gielen & De Wever, 2012, 2015).

RQ2: What is the impact of helping assessees to filter out the feedback they receive on their own PF skills as assessors?

(H2a) Students in the experimental condition will provide PF of higher quality as assessors at an earlier occasion. Therefore, it is expected to find more negative verifications and suggestive elaborations in the experimental group. (H2b) Additionally, it is expected to find an interaction effect between the PA practice (RQ1) and the experimental condition (RQ2) resulting in a quality increase during the later feedback occasions after students have experienced the filter-out scaffold.

RQ3: Do the perceived PA skills change over time? Are they related to the actual change in the PF quality?

(H3) It is expected to find an increase of PA skills based on the PA practice (PA skills will increase over time) and the effect of the experimental manipulation (the participants in the experimental condition will perceive a faster increase of their skills).

RQ4: Do students perceive the PF as useful?

(H4) It is expected to find an increase as students will have multiple occasions for PA practice. Due to the active scaffold of filtering out the received PF, it is expected that students in the experimental condition report a higher level of perceived usefulness.

Method

Participants

Participants in this study were 36 11th grade secondary students (MAge = 15, Range = 14-16) equally spread over two classes, with two different teachers. The majority was female (80.6%). All students were enrolled in the theory-oriented general secondary education track and had no prior experience with the specific PA task (i.e. assessment of a group product).

Procedure

Students worked in small groups (12 groups) on a topic which they chose concerning a specific internship institute (e.g. a local library). The learning goal was to experience the valuableness of conducting research and explore the necessary skills. After having received an introduction lecture on this topic, the groups designed a research proposal and conducted it during their internship (e.g. analyzing costumers' buy intentions in a recycling store). In the first semester the groups presented the research proposal and, at the end of the second semester, they presented the results. Both presentations were done in front of the classroom group, and each presentation was assessed by their classmates. Assessors were told that their PF would not affect their course grade to avoid possible worries about the effects of PF.

The function of the PA activity was formative in nature as the teachers intended to let the students learn things from their peers' feedback. However, to stimulate effort and justify the investment of time in the presentation, the mean PA score of the group presentation was taken into account for 15% of the course grade.

Regarding the PA scaffolds, all the participants were involved in the selection of the assessment criteria: the teacher provided a rubric that was discussed in the classroom and changes were incorporated when needed. For example, for the presentation-related criterion *coherence between speakers* level one was proposed as *Speakers weren't aware about each others' content*. Students added a part they felt was missing: *Speakers weren't aware about each others' content, which resulted in the same content being told twice.* The final rubric (see Appendix 1) had six criterions, each with five quality levels. Out of the 6 criterions, three were presentation-related,

and the other three were content-related. Since the task was different from the first group presentation (introducing the research) and the second group presentation (presenting the results of the research) the three criterions changed. Additionally, the assessors in both conditions received three guidelines to support them while giving feedback: (1) make sure your feedback is specific and linked to the matching rubric criteria, (2) give suggestions for a future improved performance, and (3) appoint the strengths, but don't be hesitant to indicate weaknesses. Finally, in the experimental group students were asked to filter out the feedback they received via three guiding questions: What feedback from your peers do you take along in preparation of a) the research project that you will conduct? b) the presentation of the research results (semester 2)? c) Please formulate for each group member a strength and an aspect that needs improvement, based on the input of your peers (see Appendix 2 for an example of the form). As a means of a manipulation check, students in the experimental condition were asked to show and shortly elucidate the completed filter-out file to the teacher after they had received the FB of their first presentation.

Regarding the PA procedure, each student acted 10 times as an assessor within his/her class (i.e. 12 sessions per class) and 2 times as an assessee. As mentioned earlier, the identity of the assessor remained anonymous. This was facilitated through the use of Mobile Response Technology (MRT), in which assessors get the opportunity to give immediate anonymous PA scores (quantitative part) and PF (qualitative part) via web-enabled devices such as smartphones, tablets or laptops (Magaña & Marzano, 2014). In this study the free tool Socrative™ (Bèta Release) was used. Every PA session included three steps as depicted in Figure 1. After all the assessors evaluated the presenting group, the results were projected and verbally discussed in the classroom. The teacher moderated this discussion phase by asking reflective questions (e.g. what is the reason for the high number of remarks on the presentation structure?). Additionally, the Socrative reports (automatically generated anonymized Excel files) were sent to the assessed group. It is important to mention that the teacher had the possibility to identify the assessors in case of unfriendly or hostile messages were given.

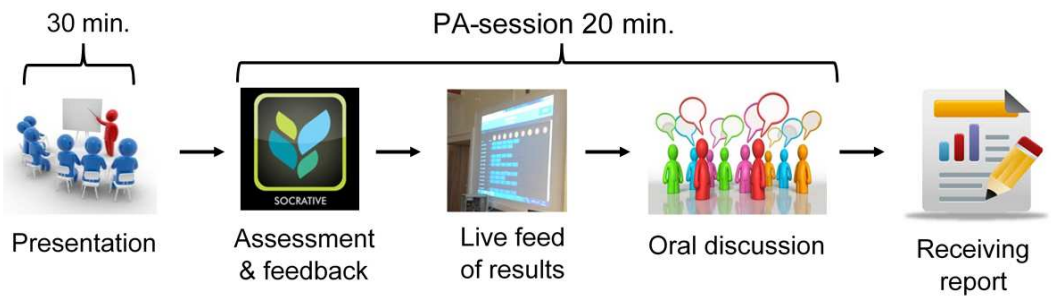


Figure 1. PA session

Measurements

The content analysis (RQ 1 & 2)

To measure the evolution of PF quality over time, the feedback content was analyzed at individual level at three occasions (henceforth FB Occasion 1, FB Occasion 2 and FB Occasion 3) on a subsample of 20 of a total of 24 sessions (6 sessions in December, 2 sessions beginning of June and 2 at the end of June per group). See Figure 2. This resulted in a database of 1561 segments. The first two levels (i.e. PF style and PF type) of the hierarchical content-analysis scheme by Gielen and De Wever (2015) were used with a slight modification. PF style here includes two categories verification and elaboration; while the third category from Gielen and De Wever –general – was not included as it was not identified among our data. Regarding PF type, there are five categories: positive verification, neutral verification, negative verification, informative elaboration and suggestive elaboration (see Table 1). As only a very small amount of neutral verifications was found and only during FB Occasion 2, these were not presented in the analyses. Additionally, we added another level to our data: whether the PF related to a content-related criterion in the rubric or to a presentation-related criterion.

Data were coded by the first author and an external coder that was trained for the task. A random subsample containing 38.41% of the total segments was coded by both with a Krippendorff's Alpha of .99 for the content-related criteria, and .98 for the presentation-related criteria. Next, 7 out of the 20 feedback sessions were double coded resulting in 600 segments (267 content-related / 333 presentation-related).

The alpha values were above or equal to the popular benchmark of .80 (De Swert, 2012; Landis & Koch, 1977): content-related PF style (.88), presentation-related PF style (.98), verification type (.97) and elaboration type (.97).

Students' PF skills perception (RQ3)

Participants reported their PF capability using a 10-point slider scale (0 totally not capable – 10 totally capable; rounded to 1 decimal place), in 3 items (example item: *Rate your capability of being able to formulate suggestions for improvement regarding a peers' work*). This scale was measured before the start of the intervention ($\alpha = .79$), after the session in semester I ($\alpha = .88$), and after all sessions in semester II ($\alpha = .94$).

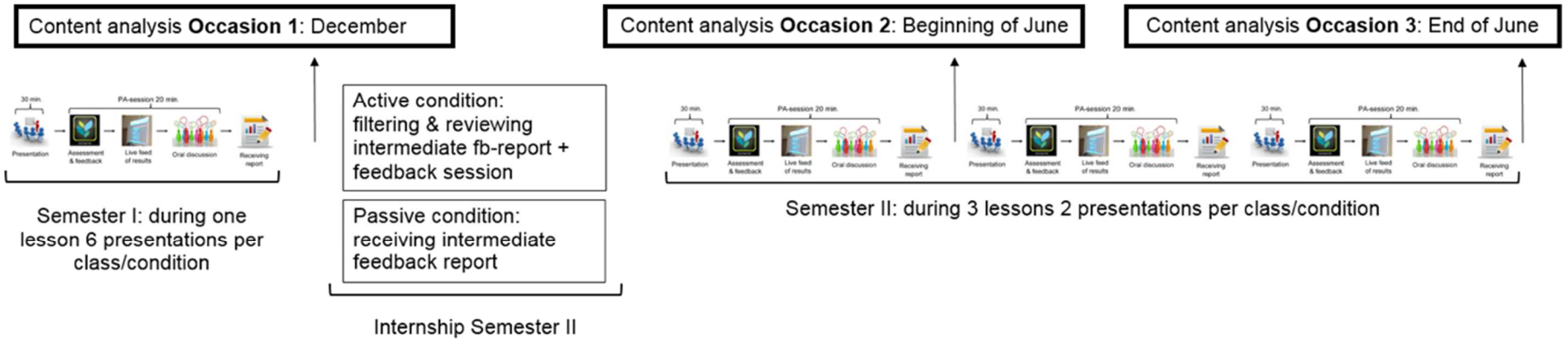


Figure 2. Content analysis FB occasions

Students' perceived usefulness of the received PF (RQ4)

This variable was measured through a 3-items 7-points Likert-scale (example item: *The feedback in the Socrative report was useful for future presentations).* Reliability analysis showed acceptable scores ($\alpha^{\text{Semester I}}$ = .69 / $\alpha^{\text{Semester II}}$ = .67). Furthermore, an open-ended question was posed on this issue: *Do you think the feedback in the Socrative report was useful or not? Please, explain why (not).*

Table 1
Coding scheme for analysing PF content quality (modification based on Gielen et al., 2015)

Category	Subcategory	Description	Examples
PF style	Verification	The feedback segment is an evaluative statement expressed as a positive, neutral or negative remark on past performance.	Content-related: *You gave a good explanation of which steps you're planning to take.* Presentation-related: *The switch between the speakers was smooth and without interruptions.*
	Elaboration	The feedback segment is an informative statement that builds further on verification or remark expressed as e.g. a question, a confirmation, a suggestion or a justification.	Content-related: *[…] because I think it will be difficult for such young children to fill up a questionnaire.* Presentation-related: *Try to look a bit more at the public.*
Verification type	Positive	The feedback segment is a positive evaluative statement.	Content-related: *The research methods were well chosen.* Presentation-related: *She speaks relaxed.*
	Negative	The feedback segment is a negative evaluative statement.	Content-related*: The connection between the research question and the conclusion was not that clear.* Presentation-related: *A lot of content was just read out loud.*
	Neutral	The feedback segment is a neutral evaluative statement.	Content-related: *It is a pity that your response rate was that low.* Presentation-related: *All dots were consecutively presented on a slide.*
Elaboration type	Informative	The feedback segment is an informative statement, which gives more details about a previous evaluative statement without activating the student to adapt his work.	Content-related*: Good that you used graphics (pos. verification),* ***that makes it more understandable.*** Presentation-related: *I thought the PPT was confusing (neg. verification),* ***maybe this was due to the particular theme.***
	Suggestive	The feedback segment is a suggestive statement, which gives more details about a previous evaluative statement with the purpose to activating the student to adapt his work.	Content-related: *For future presentations, try to give some more explanation on the categories in the diagrams.* Presentation-related: *Try to divide the texts fragments more amongst all participating speakers.*

Data analysis

As mentioned earlier, the qualitative content data was treated quantitatively. Repeated measures ANOVAs were performed for all content categories with estimable amounts of feedback messages. The mean number of segments per assessor per session of a specific category was entered as a dependent variable, and condition as between subjects variable.

Results

RQ1: What is the evolution in PF quality over time when students practice PA several times?

First, it is important to mention that the overall quality of the presentations was comparable. The mean rubric score was 3.83 on 5-point Likert scale (*SD* = .30, Min = 3.25, Max = 4.23). In line with our hypothesis (H1) the effect of PA practice increased significantly the number of negative verifications in the presentation-related criteria category over time [$F(2, 68) = 2.68$, $p = .041$, $\eta_G^2 = .06$] (see Table 2). This means that after multiple sessions students gradually dare to indicate more weaknesses in a peer's work regarding presentation-related aspects. Additionally, contrast analyses revealed that between FB Occasion 1 and 3 [$F(1, 34) = 5.211$, $p = .029$, $r = .365$] there were significantly more negative verifications given. For the content-related negative verifications, there was no effect of time [$F(2, 68) = .41$, $p = .669$]. In relationship to the frequency of *positive verifications* there was a significant evolution: the number of content-related positive verifications augmented among the different FB Occasions [$F(2, 68) = 9.48$, $p = .000$, $\eta_G^2 = .17$]: Occasion 1 and 2 [$F(1, 34) = 14.102$, $p = .001$, $r = .541$]; Occasion 1 and 3 [$F(1, 34) = 13.984$, $p = .001$, $r = .539$]. Regarding the presentation-related criteria no significant effect was found [$F(2, 68) = 2.326$, $p = .105$].

Table 2.
Verification type: descriptives, mean amount of positive and negative verifications per student per session for content- and presentation-related criteria

Verification type		Occasion 1 M*(SD)*	Occasion 2 M*(SD)*	Occasion 3 M*(SD)*
Content-related criteria	Positive	.94*(.50)*^{u}	.56*(.61)*^{u,v}	1.10*(.78)*^{v}
	Negative	.29*(.29)*	.24*(.37)*	.31*(.47)*
Presentation-related criteria	Positive	1.01*(.47)*	1.21*(.81)*	1.42*(1.05)*
	Negative	.37*(.27)*^{w}	.45*(.54)*	*.63(.71)*^{w}

Note: same superscript u,v,w indicate significant differences at $p < .05$

Regarding elaboration (Table 3), informative elaborations of content-related criteria increased [F(2, 68) = 5.524, p =.006, η_G^2 = .115]. Contrast analyses revealed a significant increase between FB Occasion 1 and FB Occasion 2 [F(1, 34) = 4.155, p =.049, r = .329] and FB Occasion 1 and FB Occasion 3 [F(1, 34) = 10.019, p =.003, r = .477]. Additionally for the informative elaborations in the PF messages on presentation-related criteria, a likewise increase was found: there was a significant main effect of Time [F(1.55, 52.78) = 5.693, p =.01, η_G^2 = .108] while applying a Greenhouse-Geisser correction. Contrast analyses revealed a significant difference between FB Occasion 1 and FB Occasion 2 [F(1, 34) = 4.672, p =.038, r = .348] and FB Occasion 1 and FB Occasion 3 [F(1, 34) = 8.175, p =.007, r = .440]. Overall, we can say that students add more elaborative information in their PF messages when they get multiple practice opportunities.

As expected, there was a significant effect of practice in the suggestive elaborations in presentation-related PF messages, [F(2,68) = 5.875, p = .004, η_G^2 = .131]. Contrast analyses revealed a significant increase of suggestive elaborations between FB Occasion 1 and FB Occasion 3 [F(1, 34) = 12.982, p = .001, r = .524].. Regarding *suggestive elaborations* for the content-related criteria there was no significant main effect of practice [F(2,68) = 1.491, p = .232].

Table 3
Elaboration type: descriptives, mean amount of informative and suggestive elaborations per student per session for content- and presentation-related criteria

Elaboration type		Occasion 1 M*(SD)*	Occasion 2 M*(SD)*	Occasion 3 M*(SD)*
Content-related criteria	Informative	.17*(.20)*^{u,v}	.32*(.40)*^{u}	.47*(.56)*^{v}
	Suggestive	.14*(.18)*	.22*(.32)*	.14*(.28)*
Presentation-related criteria	Informative	.31*(.36)*^{w,x}	.53*(.63)*^{w}	.79*(1.01)*^{x}
	Suggestive	.16*(.19)*^{y}	.42*(.85)*	.79*(1.01)*^{y}

Note: same superscript u,v,w,x,y indicate significant differences at p < .05

In sum, it was expected that negative verifications and suggestive elaborations would increase as it was the case, but additionally informative elaborations also increased. This is actually an unexpected positive result as this adds to the previous evidence that through PA practice students improve the quality of the feedback they give as PA assessors.

RQ2: What is the impact of helping assessees to filter out the feedback they receive on their own PF skills as assessors?

Two hypotheses were tested here. First (H2a), it was explored whether the intervention in the experimental group (helping assessees to filter out information) would improve the PF quality they would give in subsequent tasks as assessors. This hypothesis has to be rejected as the intervention did not improve the PF quality in terms of negative verifications and suggestive elaborations, nor in other categories (see Table 4). Regarding H2b, it has to be rejected too as there was no effect in the interaction between practice and the experimental manipulation.

Table 4
Test statistics for main and interaction effect RQ2

		Main effect	Interaction effect
Verification type			
Content-related criteria	Positive	F(1, 34) = .546, *p* =.465	F(2, 68) = 3.133, *p* =.052
	Negative	F(1, 34) = .148, *p* =.703	F(2, 68) = 1.278, *p* =.285
Presentation-related criteria	Positive	F(1, 34) = .037, *p* =.849	F(2, 68) = 2.021, *p* =.141
	Negative	F(1, 34) = .002, *p* =.961	F(2, 68) = .633, *p* =.534
Elaboration type			
Content-related criteria	Informative	F(1, 34) = .224, *p* =.639	F(2, 68) = .577, *p* =.544
	Suggestive	F(1, 34) = .619, *p* =.437	F(2, 68) = .005, *p* =.995
Presentation-related criteria	Informative	F(1, 34) = .524, *p* =.472	F(1.55, 52.78) = 1.981, *p* =.157
	Suggestive	F(1, 34) = 1.693, *p* =.202	F(2, 68) =.749, *p* =.477

RQ 3: Students' PF skills perception

When assessing students' PF skills perception, during and after the PA-sessions, a repeated measures analysis indicates a significant main effect of practice [F(2,70) = 7.64, p = .001, η_G^2 = .136]. Contrast analyses revealed a significant increase between FB Occasion 1 and 2 [F(1,35) = 10.32, p = .003, r = *.477*] and FB Occasion 1 and 3 [F(1,35) = 13.50, p = .001, r = *.528*] (see Table 5). No significant effects where found for the interaction between practice and the experimental manipulation [F(2,70) = .65, p = .53] nor for the experimental manipulation [F(1,35) = 2.50, p = .12]. Therefore, the H3 can only be partially supported as there was an effect of practice but not of the experimental manipulation (i.e. filter out guiding questions).

Additionally, it was important to check whether the students' perceptions about PF skills increase was veridical. For that reason the concurrence between PF skills perception increase and the development of PF skills was explored. As students reported higher PF skills and the content analysis indicated an improvement of negative verifications, informative elaborations and suggestive elaborations over time, this suggests that students did not only improve their PF skills

throughout the PA sessions as shown through content analysis, but that they were also aware of the learning progress they make regarding their feedback skills.

Table 5
Descriptives students' PF skills perception

Control condition			Experimental condition		
Occasion 1 M*(SD)*	Occasion 2 M*(SD)*	Occasion 3 M*(SD)*	Occasion 1 M*(SD)*	Occasion 2 M*(SD)*	Occasion 3 M*(SD)*
5.62*(1.35)*	6.24*(1.50)*	6.85*(1.23)*	6.19*(1.37)*	7.09*(1.67)*	7.07*(1.49)*

RQ 4: Students' perceived usefulness of PF

Students' perceived usefulness of the PF was highly positively evaluated after both their presentation in the first ($M_{Experimental}$ = 5.80; SD = .84 / $M_{Control}$ = 5.70; SD = .62) and second ($M_{Experimental}$ = 5.43; SD = .75 / $M_{Control}$ = 5.33; SD = .81) semester on a 7-point Likert-scale. Contrary to our H4 there is a significant decrease over time [$F(1, 33) = 5.36$, $p = .03$, $\eta_G^2 = .083$]. Despite the significant decrease, the absolute values on the 7-point Likert-scale remain highly positive. Additionally, no significant difference could be found between the two conditions [$F(1, 33) = 0.00$, $p = .99$].

Results from the open-ended question about the usefulness of the feedback in the Socrative report were very positive: 88.88% (semester I) and 94.44% (semester II) for the experimental condition and 100% (semester I) and 94.44% (semester II) for the control condition. Additionally, students stated that the plurality of opinions they received is the biggest advantage of this PA procedure.

Discussion

PF quality in PA settings is crucial for students to improve their work (Topping, 1998). Although previous research has shown that practice is central for the development of judgmental skills, literature on what exactly is the effect of PA practice on PF quality was not conducted yet. The main aim of this study was twofold: exploring the effect of both PA practice and PA scaffolds – especially "filter out" – on PF quality. Additionally, it was explored whether students' perceived an improvement of their PF skills, which was matched with their actual evolution. Finally, we explored whether the PF is perceived as useful from an assessees' point of view. The setting was a synchronous (i.e. immediate feedback) PA environment in secondary education where assessors were anonymous.

Regarding, our first hypothesis (H1) it can be maintained that PA practice has an impact on PF quality over time. The number of positive verifications remained stable, which confirms the finding of Gielen and De Wever (2015) that students usually tend to give positive verifications in their PF. This is not problematic as long as the other PF components are also present and, especially, if the students are not giving positive verifications to a particular piece that is incorrect. Aligned with the research by Gielen and De Wever (2015), (a) our participants showed more negative verifications, and (b) more suggestive elaborations. However, next to that, our research is the first one to find an increase in informative elaborations. The fact that informational feedback was beneficial for our participants performance is important because it means that, at the end of our intervention, students were able to include the three utmost difficult feedback components in their immediate feedback messages: (1) pointing out when the assessee's performance is not at the requested criteria level, (2) why this is the case and (3) how it should be improved (Hattie & Timperley, 2007). Our findings confirm the importance of practice in PA settings since students' expertise for making valuable judgments on a peers' work improves over time (e.g. Boud & Soler, 2015; Panadero et al., 2016, Sluijsmans, 2002). However, the explanation of PF quality improvement as a practice effect might be entangled with a memory effect and a norm development effect over time. More specifically, the use of criteria might have had a substantive positive effect on the quality improvement of the PF over time as students may increasingly well remember the feedback criteria. Furthermore, as also suggested by Cheng (2015), the use of feedback criteria itself could function as an incentive to make structured comments and may reinforce the PF quality. The repeated use of the rubric criteria might thus have led to some sort of 'norm development' in the classroom. Future research should focus on disentangling the aforementioned interpretations.

Although it was expected (H2a) that in the experimental condition (i.e. helping assessees to filter out useful feedback) assessors would provide PF of higher quality, the hypothesis needs to be rejected. Furthermore, hypothesis 2b cannot be maintained as no interaction effect between PA practice and the experimental condition was found. These non-significant effects might be explained by the fact that the effect of filtering out might have not been strong enough on its own and that its effect might have been diluted by the other PA scaffolds. For example, the fact that students in both conditions received a Socrative feedback report could have been a sufficient support for an adequate feedback filtering, resulting in an increase of PF quality in the consecutive PA sessions regardless of the filter out effect. In that sense, the filter-out activity for the assessees could possibly be seen as an example of an over-scripting activity, at least within this specific context and task (Dillenbourg, 2002). Another possible interpretation for the non-significant effect might have been the fact that the filter-out scaffold was given while groups were processing the received PF, and that students' PF quality was measured during those situations that students

were providing feedback. Thus, the scaffold and the PF quality in subsequent feedback provisions might have been too loosely connected in time and as a consequence this would account for the missing effect.

The findings regarding students' perceived evolution in PF skills (H3) are in line with the findings of their actual improvement over time. This finding suggests that all students acknowledged that their involvement in multiple PA sessions lead to an improvement of their PF skills. This is an important finding as identifying oneself as an active learner is considered to be a key element in the development of PF skills (Boud & Soler, 2015). Through providing and receiving PF multiple times, students are constructing meaningful feedback conceptions for themselves. Previous research has shown that this ensues several benefits such as giving students more control over the feedback processes, and as a result of this, also more control over their own learning (Nicol, Thomson, & Breslin, 2014). However, as students' perceived PF skill was measured through a quantitative sliderscale compared to the detailed qualitative analysis of the actual FB, future research might also include a more general PF quality score (for example the Peer Feedback Quality Index by Prins, Sluijsmans and Kirschner (2006)) as both measures would then be comparable through a correlation of the actual gain and students' perceived gain over time.

The assessees' perceived usefulness of the received PF (RQ4) was also analyzed, as the willingness to follow assessors' advice is essential to augment the quality of the performance. As expected (H4), the results show that the received PF was highly positively appraised by students in both conditions. The small but significant decrease in appreciation of the PF over time might be related to the fact that students did not get the opportunity to tackle the suggestions for improvement after the second PA session in semester II, which was the case in semester I. Another reason, in line with the self-determination theory by Ryan and Deci (2000), may have been the fact that students perceived the PF as something that controlled them, in a way that it hindered their self-perceived autonomy. Nevertheless, this should not be perceived as problematic since the overall score remained high. The fact that the PF was perceived useful is valuable as it is generally too easily assumed that students automatically perceive feedback as being useful (Harks, Rakoczy, Hattie, Besser, & Klieme, 2013). Furthermore, in a recent theoretical model on PA, sound feedback reception is seen as essential, because feedback helps students to form a more objective lens for self-assessment and self-regulatory processes (Reinholz, 2015). Therefore, we suggest that future research should study the impact of multiple enactments as peer assessors on students' self-regulating skills more thoroughly (Panadero et al., 2016).

Given the sample size and gender bias of the sample (mostly female), the findings of this study should be interpreted with caution. Studies with bigger sample sizes, within other settings

(e.g. higher education) and a variety of courses should be conducted in order to replicate our results. In line with the recent work by Reinholz (2015) more sustained training might be needed to stimulate even more feedback improvement. Furthermore, the impact of the PF received – although it was perceived useful - on the actual performance was not explored in this study, as advocated by Evans (2013). Future research should focus on this issue by defining quality categories for the task on which PA is performed, for example based on tasks of previous student cohorts and in close collaboration with the involved teachers. This would offer the opportunity to get insight into the impact of the different peer feedback styles.

Our findings are important for educational research and practice. Our study reveals that students in a PA setting improve the quality of their PF over time, and therefore practice should be a major component in PA implementations. As discussed by Panadero, Jonsson & Strijbos (2016) additionally to assuring such practice, teachers need to monitor the PA process and coach the students even providing feedback about the PA itself. Another important implication of our results, is that PF quality is not only mentioning if something is correct or not (i.e. positive and negative verifications), but also offer information on why this was (in)correct (i.e. informative elaborations), in combination with suggestions to improve the presentations (i.e. suggestive elaborations). This is absolutely in line with previous research on how to give adequate feedback to promote learning (Hattie & Gan, 2011). Concerning possible practice constraints of implementing multiple PA sessions, MRT has proven to be adequate to facilitate the reciprocal feedback processes, so teachers are encouraged to use this tool to organize PA practice within their classrooms.

In conclusion, our study clearly shows that when students are offered PA practice opportunities in combination with rubrics and guiding questions for the assessors (not the assessees), the more likely students will develop expertise for making sound evaluative judgements on peers' work. More specifically, content analysis of the PF messages revealed that students not only inform their peers about what is wrong and why but also provide suggestions on how to improve the performance. In sum, this study clearly indicates that PA practice in combination with clearly defined assessment scaffolds constitutes a valuable classroom assessment practice since students experience a tangible educational value of PA through the perceived and actual growth of their PF skills.

References

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research*, *61*(2), 213–238. doi:10.3102/00346543061002213

Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151–167. doi:10.1080/713695728

Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 1–14. doi:10.1080/02602938.2015.1018133

Cheng, K.-H., Liang, J.-C., & Tsai, C.-C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education*, *25*, 78–84. doi:10.1016/j.iheduc.2015.02.001

De Swert, K. (2012). Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha. Retrieved from http://www.polcomm.org/wp-content/uploads/ICR01022012.pdf

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. Heerlen, Open Universiteit Nederland. Retrieved from https://telearn.archives-ouvertes.fr/hal-00190230/

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research*, *83*, 70–120.

Falchikov, N., & Goldfinch, J. (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, *70*(3), 287–322. doi:10.3102/00346543070003287

Gielen, M., & De Wever, B. (2012). Peer Assessment in a Wiki: Product Improvement, Students' Learning And Perception Regarding Peer Feedback. *Procedia - Social and Behavioral Sciences*, *69*, 585–594. doi:10.1016/j.sbspro.2012.11.450

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, *52*, 315–325. doi:10.1016/j.chb.2015.06.019

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2013). The effects of feedback on achievement, interest and self-evaluation: the role of feedback's perceived usefulness. *Educational Psychology*, 1–22.

Hattie, J., & Gan, M. (2011). Instruction Based on Feedback. In *Handbook of Research on Learning and Instruction* (pp. 249–270). Routledge. doi:10.4324/9780203839089.ch13

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. doi:10.3102/003465430298487

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, *20*, 344–348. doi:10.1016/j.learninstruc.2009.08.005

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. doi:http://doi.org/10.2307/2529310

Lin, S. s. j., Liu, E. z. f., & Yuan, S. m. (2001). Web-based peer assessment: feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, *17*(4), 420–432. doi:10.1046/j.0266-4909.2001.00198.x

Liu, N.-F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, *11*, 279–290.

Liu, X., & Li, L. (2013). Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment. *Assessment & Evaluation in Higher Education*, *39*(3), 1–18. doi:10.1080/02602938.2013.823540

Magaña, S., & Marzano, R. J. (2014). Using Polling Technologies to Close Feedback Gaps. *Educational Leadership*, 82–83.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J.M. Spector, M. D. Merril, J. van Merriënboer, M.P. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology.* pp. 124–143. New York: Lawrence Erdbaum Associates.

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: how different types of peer feedback affect writing performance. *Instructional Science*, *37*, 375–401.

Nicol, D. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, *35*(5), 501–517. doi:10.1080/02602931003786559

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102–122. doi:10.1080/02602938.2013.795518

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Human factors and social conditions of assessment. New York: Routledge* (pp. 1–39). New York, NY: Routledge.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, *9*, 129–144. doi:10.1016/j.edurev.2013.01.002

Panadero, E., Jonsson, A., & Strijbos, J. W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault & L. Allal (Eds.), Assessment for Learning: Meeting the challenge of implementation (pp. 311-326). New York: Springer.

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, *39*(4), 195–203. doi:10.1016/j.stueduc.2013.10.005

Prins, F. J., Sluijsmans, D. M. A., & Kirschner, P.A. (2006). Feedback for general practitioners in training: Quality, styles, and preferences. *Advances in Health Sciences Education*, *11*(3), 289–303.

Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. (2010). Formative peer assessment in a CSCL environment: a case study. *Assessment & Evaluation in Higher Education*, *30*(4), 417–444. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/02602930500099219

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, *35*, 435–448. doi:10.1080/02602930902862859

Reinholz, D. (2015). Peer conferences in calculus: the impact of systematic training. *Assessment & Evaluation in Higher Education*, 1-17. doi:10.1080/02602938.2015.1077197.

Reinholz, D. L. (2015). Peer-Assisted Reflection: A Design-Based Intervention for Improving Success in Calculus. International Journal of Research in Undergraduate Mathematics Education, 1(2), 234–267. doi:10.1007/s40753-015-0005-y

Reinholz, D. (2015). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 1–15. doi:10.1080/02602938.2015.1008982

Ryan, R. M., & Deci, E. L. (2000). Self-determination Theory and the facilitation of intrinsic motivaton, social development and well-being. *American Psychologist*, *55*(1), 68–78. doi:10.1037/0003-066X.55.1.68

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, *35*, 535–550.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*, 153–189.

Sluijsmans, D. M. A. (2002). Student involvment in assessment: The training of peer assessment skills. (Unpublished doctoral dissertation). Open University of the Netherlands, Heerlen.

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Martens, R. L. (2004). Training teachers in peer-assessment skills: effects on performance and perceptions. *Innovations in Education and Teaching International*, *41*, 59–78. doi:10.1080/1470329032000172720

Topping, K. J. (1998). Peer Assessment between Students in Colleges and Universities. *Review of Educational Research*, *68*(3), 249. doi:10.2307/1170598

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), (pp. 55–87). Dordrecht, The Netherlands: Kluwer Academic.

Topping, K. J. (2009). Peer Assessment. *Theory Into Practice*, *48*(1), 20–27. doi:10.1080/00405840802577569

Tsivitanidou, O. E., & Constantinou, C. P. (2016). A study of students' heuristics and strategy patterns in web-based reciprocal peer assessment for science learning. *The Internet and Higher Education*, *29*, 12–22. doi:10.1016/j.iheduc.2015.11.002

Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction*, *21*(4), 506–519. doi:10.1016/j.learninstruc.2010.08.002

Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, *20*, 270–279.

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, *81*, 123–132. doi:10.1016/j.compedu.2014.10.001

Yu, F.-Y., & Sung, S. (2015). A mixed methods approach to the assessor's targeting behavior during online peer assessment: effects of anonymity and underlying reasons. *Interactive Learning Environments*, 1–18. doi:10.1080/10494820.2015.1041405

Appendix 1

Evaluation criteria used during Semester I

Presentation-related criteria
8. Structure and lay-out
9. Contact with the audience
10. Coherence between speakers
Content-related criteria:
11. Delineation of research question
12. Methodology
13. Alacrity of participants

Example A: Coherence between speakers

Level 1	Level 2	Level 3	Level 4	Level 5
The speakers weren't aware about each other's content, which resulted in the same content being told twice.	The speakers weren't well prepared which resulted in an in-coherent presentation.	The speakers took turns in presenting, although the transition was sometimes abrupt.	The speakers took turns in a smooth way.	The speakers were well-prepared and presented in a coherent way.

Evaluation criteria used during Semester II

Presentation-related criteria
1. Structure and lay-out
2. Contact with the audience
3. Coherence between speakers
Content-related criteria
4. Explanation of different steps taken in the study
5. Presentation of the results
6. Formulation of conclusions

Example B: Formulation of conclusions.

Level 1	Level 2	Level 3	Level 4	Level 5
Based on the results the group presents wrong conclusions.	The group changed the original research questions and thus formulates inconsequent conclusions.	The group doesn't give an answer to some of the research questions.	The group formulates conclusions that are too generalized for the studied sample.	The group presents concise and clear conclusions.

Appendix 2

Dear students,
Please review the Socrative report that you received and answer following questions:

1. What feedback from your peers do you take along in preparation of…
    a. the research project that you will conduct?

    b. the presentation of the research results (semester 2)?

2. Please formulate for each group member a strength (+) and an aspect that needs improvement (-). Feel free to discuss them with your fellow group members:
    a. Group member 1 : ……

+

-

    b. Group member 2 : ……

+

-

    c. Group member 3 : ……

+

-

6 Anonymity as an instructional scaffold in PA: its effect on PF Quality and evolution in students' perceptions about PA skills

This chapter is based on:

Rotsaert, T., Panadero, E., & Schellens, T. (submitted). Anonymity as an instructional scaffold in peer assessment: its effects on peer feedback quality and evolutions in students' perceptions about peer assessment skills. Manuscript under minor revision for publication in a forthcoming special issue in the *European Journal of Psychology of Education.*

Chapter 6
Anonymity as an instructional scaffold in PA: its effects on PF quality and evolution in students' perceptions about PA skills

Abstract

Although previous research has indicated that providing anonymity is effective to create a safe peer assessment (PA) setting, continuously ensuring anonymity prevents students from experiencing genuine two-way interactive feedback dialogues. The present study investigates how fading anonymity over time can overcome this problem. 46 Bachelor students in Educational Studies participated in multiple PA cycles in which groups of students assessed each other's work. Both students' evolution in peer feedback (PF) quality as well as their perceptions were measured. The content analysis of the PF messages revealed that the quality of the PF increases in the anonymous phase and that over time the FB quality in the consecutive non-anonymous sessions was of similar quality. Results indicate that the fading approach does not hinder the perceived growth in PF skills, nor does it have a negative impact on their general conceptions towards PA. Furthermore, students clearly differentiate between their attributed importance of anonymity and their view on the usefulness of a fading approach. Findings suggest that anonymity can be used as a valuable scaffold to ease students' importance level towards anonymity and their associated need for practice.

Introduction

Peer Assessment (PA) is an assessment method in which students are actively involved in assessment processes. In PA activities students are required to reflect on the quality of a peers' work and discuss how well it corresponds with the explicitly stated goals or criteria of the work (Strijbos & Sluijsmans, 2010).

Peer Feedback (PF) is a central component of PA being the information that a student provides to a peer (Topping, 1998). Students are engaged in high-level cognitive processing during the PF process as it requires skills such as explaining, identifying mistakes and gaps and providing suggestions for improvement (King, 2002). For this reason, feedback is seen as an essential factor of the learning process (Hattie & Timperley, 2007). Furthermore, PF exemplifies

professional practice in which colleagues collaborate and offer input to improve current work (Van der Pol, Van den Berg, Admiraal, & Simons, 2008).

However, as PA is a fundamentally social and collaborative learning activity, learners' interpersonal beliefs can negatively impact its outcomes (Panadero, 2016; Raes, Vanderhoven, & Schellens, 2013). This includes possible reciprocity effects and negative feelings caused by interpersonal processes such as friendship marking due to friendship bounds, psychological unsafety, fear of disapproval when giving a low score or critical feedback (i.e. recrimination) and distrust in own and others' evaluative capabilities (e.g. Harris & Brown, 2013; van Gennip, Segers, & Tillema, 2010; Vanderhoven, Raes, Montrieux, Rotsaert, & Schellens, 2015). The recognition of the presence of interpersonal processes in PA is thus important, because the pressure students experience in the process in turn may directly impact how they view the value of PA (Li, 2016).

Previous research has shown that providing anonymity to the assessors can help to relieve the interpersonal burden from students, especially in PA activities where a score is given (Yu & Liu, 2009). Providing anonymity leads to more positive conceptions towards PA (Vanderhoven et al., 2015), helps students to become more willing to give critical feedback (Howard, Barrett, & Frick, 2010) and aims to assure accuracy (Topping, 2010). However, anonymous settings do not reflect daily face-to-face situations in which people give and receive feedback with known identities.

Accordingly, in order to involve students in reflective criticism of the products of their peers, classroom interventions are needed that a) recognize PF as an essential component of PA, b) acknowledge the inherent social nature of the PA process, and c) guide students towards an open dialogic non-anonymous feedback environment in which they can develop sustainable assessment skills (Boud & Soler, 2015; Carless, Salter, Yang, & Lam, 2011).

Peer Feedback Quality

The process of assessing and commenting on the strengths and weaknesses of a peer's work can help familiarize the assessors with the evaluation criteria, and thus to develop knowledge on what constitutes high-quality work (Cho & Cho, 2010). In essence, well-formulated feedback should provide an answer to three questions: 'Where am I going?' (feed up), 'How am I going?' (feedback) and 'Where to next?' (feed forward) (Hattie & Timperley, 2007). However, it cannot be expected that every student will offer high-quality feedback as it requires high-level cognitive processing (Strijbos, Narciss, & Dünnebier, 2010) as students need to be capable to deal with specific assessment criteria to judge a peer's performance (Gielen & De Wever, 2015). In this study PF quality is defined in terms of content characteristics. The advantage of this approach is that such characteristics are not task-specific, thus letting students focus on these characteristics might

result in a generic transferable evaluative expertise. More specifically, previous research indicated that qualitative feedback should contain two types of information: verifications and elaborations (Narciss, 2008). *Verification* refers to 'a dichotomous judgment to indicate that a response is right or wrong'. In other words, it tells the assessee whether a certain criterion was met or not. Gielen and De Wever (2015) found in their study in asynchronous online environments that students tend to give more positive than negative verifications. When offered a more structured environment (i.e. via a PF template) students gave more negative verifications. The negative verifications are necessary to expose shortcomings in a peers' performance. *Elaboration* refers to 'relevant information to help the learner in error correction' (Hattie & Gan,2011, p. 253). These types of information are thus seen as the structural components of feedback, because students require feedback that tells them not only if they performed the task (in)correctly (feedback), but also why and what they should do to improve their work (feedforward) (e.g. Prins, Sluijsmans, Kirschner, & Strijbos, 2010). Therefore, offering elaborations that justify the verification (correct vs. incorrect) is presumed to be beneficial for students' learning. As a consequence, a balanced proportion of verifications and elaborations is more valuable than just providing verifications alone (Gielen & De Wever, 2015). Furthermore, previous research shows that practice is crucial for the development of PA skills (Sluijsmans, 2002). The more practice in PA processes, the more likely students will develop the expertise for making sound PA judgements (Panadero, 2016). A recent study by Rotsaert, Panadero, Schellens and Raes (2017) in an anonymous synchronous face-to-face PA environment revealed that PA practice improved PF quality in terms of content characteristics: messages contained more negative verifications, and informative and suggestive elaborations after the intervention. The findings are promising because in this synchronous setting, compared to an online asynchronous environment, students get less time to formulate their feedback message (Tsai, Lin, & Yuan, 2002). However, when anonymity remains preserved in a face-to-face PA setting it does not reflect authentic feedback dialogues as, for example, in work environments (see further).

Perceived Peer Feedback Skills

The potential of learning and assessment activities depend on the way students perceive them (Boud & Soler, 2015). If students perceive they are becoming more capable as peer assessors, they will be more motivated to perform PA and believe it is useful (Author, 2015). Therefore, it is important to consider students' perceptions about the improvement of their PF skills when studying their evolution in PF quality. Additionally, the alignment of students' perceived PF skills and the actual quality of their generated PF quality is important. A recent study revealed that by providing PF multiple times, students reported having experienced an increase of their PF skills

(Rotsaert et al., 2017). Previous research has shown that this ensues several benefits such as students having more control over the feedback processes, and as a result also more control over their own learning (Nicol, Thomson, & Breslin, 2014). This might support the development of students' sustainable assessment skills, that is, increasing the capacity for judging one's own future work (Boud & Soler, 2015).

Anonymity as an instructional scaffold within PA

Literature on social psychology suggests that non-anonymous and anonymous interactions may produce differential effects on participants' perceptions toward their interaction counterparts, the interaction space and the experience itself (for a detailed discussion see Panadero 2016, and Yu & Sung, 2015). Within PA-settings, the existing research suggests that anonymity is one of several factors that encourages student participation (Ballantyne, Hughes, & Mylonas, 2002). Additionally, due to its role in diminishing reciprocity effects anonymity might result in a fairer assessment (Freeman & McKenzie, 2000). In essence, when students are interpreting assessment and feedback that they received from peers, the social context can be critical because PA "does not happen in a vacuum; rather it produces thoughts, actions, and emotions as a consequence of the interaction of assessees and assessors." (Panadero , 2016, p.2). From previous studies we know that anonymity of the assessors offers a sense of psychological safety which is defined by Yu and Sung (2015) as "a shared belief denoting one's emotional ability to take an interpersonal risk without fearing negative consequences with regard to one's well-being, self-image, and status" (p.3). Another frequently referred advantage of anonymity in PA is that it helps in decreasing ones' fear of disapproval when giving a low score or negative feedback. Furthermore, students seem to experience less peer pressure (Vanderhoven et al., 2015). Although current research tends to favour the use of anonymity, in a review study Panadero (2016) pointed out a tension between implementing anonymous PA and the formative use of PA: anonymity might support to lessen the impact of interpersonal processes, while simultaneously anonymity might also hinder the creation of a rich and interactive feedback environment. When implementing PA teachers are thus challenged to find a balance between the creation of a safe learning environment provided through anonymity and the creation of a rich PF setting, which will consequently take more class time (Panadero, 2016). To date no clear guidelines are available for teachers to cope with this tension.

Moreover, there is currently a lack of studies that focus on the actual interaction behaviour of students in anonymous PA settings. The theoretical foundations (Yu and Sung, 2015) on the possible impact of anonymity imply that students would enact different feedback behavior depending whether their identity as assessors is revealed or not. The findings of current studies

on this topic are scarce and contrasting. For example, in a series of studies by Yu and colleagues (Sung, Chang, Chang, & Yu, 2010; Yu & Liu, 2009) no significantly different effects on the actual interaction behaviour of the participants in the non-anonymous and anonymous online PA conditions were found. On the contrary, Howard and colleagues (2010) found in an online asynchronous PA setting that students who were anonymous were approximately five times more likely to provide critical feedback than were those whose identities were known to their recipients.

The relationship between attributed importance of students towards anonymity and the quality of feedback has to date not been sufficiently explored. Furthermore, it has been suggested that the relevance of anonymity should be explored in real face-to-face PA settings as current studies mostly focus on online PA-settings (Ainsworth et al., 2011). Building on the suggestions of both Howard et al. (2010) and Vanderhoven et al. (2015) who state that anonymity could function as a temporary scaffold in order to acknowledge students' possible feelings of unsafety and psychological discomfort and to use it as a facilitator to be able to generate critical feedback without social repercussions, we intend to explore this hypothesis by gradually evolving from an anonymous PA towards dialogic feedback in a face to face setting. In dialogic feedback interpretations are shared, meanings negotiated and expectations clarified between assessor and assessee (Carless, 2011). In that sense learners can practice giving feedback knowing they are not vulnerable to social repercussions (Howard, 2010). In order to be able to study this practice phase we created a face-to-face PA activity in which students experienced the transition from an anonymous to a non-anonymous setting.

Research questions & hypotheses

The aim of this study is to explore the effects of fading anonymity on PF quality while exploring students' perceptions about the process. The specific research questions and hypotheses are:

RQ1: How does PF quality change over time when students consecutively practice PA in an anonymous and non-anonymous setting?

> (H1) It is expected to find an overall increase of negative verifications and informative and suggestive elaborations because the anonymity fading effect should help students to become willing to point out/elaborate on weaknesses in peers' work and formulate suggestions for improvement.

RQ2: Do students' perceived PF skills change over time in a PA setting with a transition from anonymous to non-anonymous?

(H2) It is expected to find an increase on perceived PF skills due to students' appreciation for the transition towards non-anonymity in the PA setting.

RQ3: How does the transition from an anonymous to a non-anonymous PA affects students' perceptions regarding a) importance of anonymity, b) their perceptions towards interpersonal variables and c) their general conceptions towards PA?

(H3a) Importance of anonymity will be lower after the intervention as they will by then appreciate the authentic and personalized PA and PF.

(H3b) Regarding the positive interpersonal processes (psychological safety, trust, and value congruency) it is hypothesized to find an overall increase. Regarding the negative interpersonal variables (fear of disapproval and friendship marking) the opposite evolution is expected.

(H3c) An overall increase in students' general conceptions towards PA is expected, as students will appreciate the imposed transition between anonymity modes and its effects on their PF skills.

Method

Participants

Participants in this study were 46 third-year bachelor students in Educational Studies that were enrolled in the course Instructional Design. Their mean age was 21 years and the majority were female (84.44%). 88.6% of the participants had prior experience with PA.

Procedure

Students received a group assignment (16 groups) to prepare and present a workshop (max. 30 min.) on one of the provided topics (e.g. The Jigsaw Classroom). Students worked in small groups (16 groups). The learning goal was to organize a short workshop in a team and choose an appropriate format to transfer new learning content. They received an introduction lecture on the theoretical background of PA and on the pedagogical principles, in which examples of expert presenting performance (i.e. modelling) were shown and assessed by the students via an example rubric that was used in a previous PA project. Students were asked to apply the pedagogical principles they had learned in the lectures (e.g. principle of gradualism). In order to avoid too lengthy workshop sessions the class was divided (2 times 8 groups - called A & B).

Regarding the PA-task, students assessed their peers on the content (group level) and the presentation of the workshop (individual level) in terms of rubric scores and criteria-related feedback. Assessors were told that their PF would not affect their grade in the course to avoid possible apprehensions. The function of the PA-activity was formative in nature as the university teachers' intention was to promote that students learn from their peers' use of didactical principles and presentation styles.

Regarding the PA procedure, each student acted 7 times as an assessor (i.e. a total of 8 PA-sessions per class over 4 weeks) and 1 time as assessee. In order to be able to test our hypotheses, the feedback of the assessors remained anonymous during the first two weeks, and became known the last two weeks of the intervention. This was facilitated through the use of Mobile Response Technology (MRT), in which assessors get the opportunity to give immediate (non) anonymous PA scores and PF via web-enabled devices such as smartphones, tablets or laptops (Magaña & Marzano, 2014). In this study the free MRT-tool Socrative™ was used.

Every PA-session included three steps as depicted in Figure 1. First, all the assessors evaluated the presenting group. Second, the results (i.e. both rubric scores and feedback messages) were projected and verbally discussed in the classroom. In order to evolve to dialogic feedback, the teacher moderated this discussion phase by asking reflective questions. This included both content related input to enforce shared understanding of the criteria (e.g., What is the reason for the high number of remarks on the presentation structure?) as well as social-affective input, involving acts that build up trust and scale up mutual support between assessors and assessees (e.g., Do not worry too much when your timing was criticized, we all know it takes a lot of experience to get this right. Envision this task as a practice opportunity.) (Xu & Carless, 2016). During the first two weeks (anonymous phase) the reflective questions remain broad as such students' could answer the teachers' question, without explicitly referring to their personal

feedback message in the system. In the last two weeks (non-anonymous phase) however we aimed for two-way dialogic feedback in which there was a possibility to refer directly to students' input in the Socrative system. The role of the teacher was to facilitate the discussion on the strengths and weaknesses of the workshops. Third, the Socrative reports were sent to the assessors. Based on observations from previous likewise studies (e.g., Rotsaert et al. (2017), in both phases the teacher had the possibility to identify the assessors in case unfriendly or hostile messages were given.

In a way to implement the best possible conditions for PA to promote students' learning, two additional scaffolds were used. First, students developed their own 5-level rubric with 3 didactic principles and 3 presentation related criteria (Panadero, Romero, & Strijbos, 2013). See Appendix. Second, as suggested by Reinholz (2015), during the workshop sessions the assessors received three guidelines to support them while giving feedback: (1) make sure your feedback is specific and linked to the matching rubric criteria, (2) give suggestions for future improved performance, and (3) appoint the strengths, but don't be hesitant to indicate weaknesses.

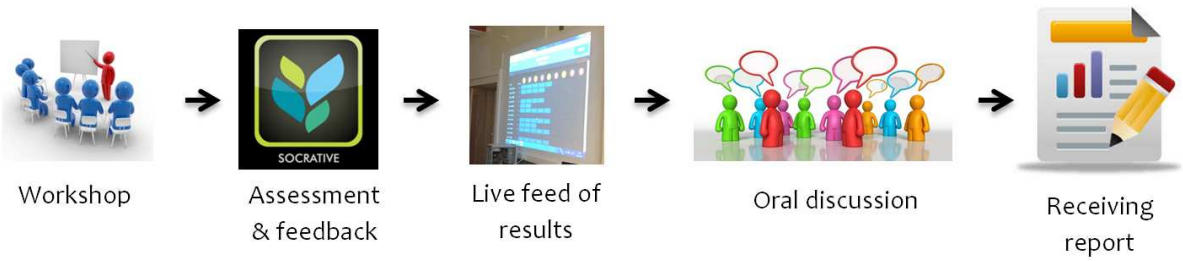


Figure 1. PA-Session

Measurements

Content analysis (RQ1)

To measure the evolution of PF quality, the feedback content was analyzed at four occasions (henceforth Session 1, Session 2, Session 3 and Session 4) on a random subsample of 8 of a total of 16 workshops (2 groups were analyzed per session). This resulted in a database of 4390 coded segments.

The first two levels (i.e. PF style and PF type) of the hierarchical content-analysis scheme by Gielen and De Wever (2015) are used with a slight modification. PF style here includes three categories verification, elaboration and general; which refers to general statements that can be labelled as neither verification nor elaboration. Regarding PF type, there are five categories: positive verification, neutral verification, negative verification, informative elaboration and suggestive elaboration (see Table 1). Additionally, as we make a distinction between two types of criteria, we added another level to our data: whether the PF related to a content-related criterion or a presentation-related criterion in the rubric.

Table 1

Coding scheme for analysing PF content quality (modification based on Gielen et al., 2015)

Category	Subcategory	Description	Examples
PF style	Verification	The feedback segment is an evaluative statement expressed as a positive, neutral or negative remark on past performance.	Content-related: *There were no concrete examples given.* Presentation-related: *The interaction was very well planned in the session.*
	Elaboration	The feedback segment is an informative statement that builds further on verification or remark expressed as e.g. a question, a confirmation, a suggestion or a justification.	Content-related: *[…] because of the drawing we could make ourselves* Presentation-related: *Try to look a bit more at the public.*
Verification type	Positive	The feedback segment is a positive evaluative statement.	Content-related: *The theory part was very well structured.* Presentation-related: *She speaks relaxed.*
	Negative	The feedback segment is a negative evaluative statement.	Content-related*: The new theory was mentioned too briefly.* Presentation-related: *A lot of content was just read out loud.*
	Neutral	The feedback segment is a neutral evaluative statement.	Content-related: *It is a pity that your response rate was that low.* Presentation-related: *The presentation had to be given in short timespan.*
Elaboration type	Informative	The feedback segment is an informative statement, which gives more details about a previous evaluative statement without activating the student to adapt his work.	Content-related*: It is a pity that you didn't show your own website (neg. verification),* ***that made it a bit boring.*** Presentation-related: *It was good that you used Prezi (pos. verification),* ***this made the whole more dynamic.***
	Suggestive	The feedback segment is a suggestive statement, which gives more details about a previous evaluative statement with the purpose to activating the student to adapt his work.	Content-related: *For future presentations, try to give some more examples.* Presentation-related: *Try to divide the texts fragments more amongst all participating speakers.*

Data was coded by the first author and an external coder that was trained for the task. For the segmentation process a random subsample of feedback messages of 5 out of the 8 coded sessions were segmented (N = 659) and double coded by both. For the segmentations process Krippendorff's Alpha was .99 for the content-related criteria, and .98 for the presentation related criteria. Additionally, the hierarchical double coding of 1977 segments resulted in alpha values which were above or equal to the popular benchmark of .80 (De Swert, 2012; Landis & Koch, 1977), with an average of .98.

Students' PF skills perception (RQ2)

Participants reported their PF capability using a 10-point slider scale (0 *totally not capable* – 10 *totally capable*; rounded to 1 decimal place), in 3 items (example item: *Rate your capability of being able to formulate suggestions for improvement regarding a peers' work*). This scale was measured before the start of the intervention (α = .88), after the first anonymous session (α = .92), after the first non-anonymous session (α = .95) and after the last session (α = .93).

Students' perceptions towards anonymity, interpersonal variables and conceptions towards PA (RQ3)

These variables were measured before the intervention (henceforth Time 1), after the anonymous (henceforth Time 2) and at the end of the non-anonymous sessions (henceforth Time 3) (see Table 2). Except *fear of disapproval* and *friendship marking* as students first needed to experience the specific PA setting before their opinion was sounded. All items were measured using a 7-point Likert scale, and anchored by 1 (*totally disagree*) and 7 (*totally agree*).

Table 2
Students' perceptions towards anonymity, interpersonal variables and conceptions towards PA

	Number of items	Cronbach's α Time 1	Cronbach's α Time 2	Cronbach's α Time 3
Importance to Anonymity Example item: I think it is important that my PF is given anonymously.	4	.85	.91	.87
General Conceptions of PA Example item: PA is useful.	6	.87	.86	.85
Trust in Own Evaluative Capabilities Example item: I have confidence in my capacity to give feedback to peers.	4	.83	.87	.89

	Number of items	Cronbach's α Time 1	Cronbach's α Time 2	Cronbach's α Time 3
Trust in Peers' Evaluative Capabilities Example item: My peers are capable of giving my feedback.	4	.68	.87	.89
Value Congruency Example item: My peers have the same expectations about the goals of this PF activity.	4	.65	.84	.81
Psychological Safety Example item: I feel comfortable giving feedback to peers in this group.	4	.90	.84	.86
Friendship marking Example item: I gave friends higher rubric scores and more positive feedback.	4	/	.92	.85
Fear of Disapproval Example item: I was afraid that when I give low rubric scores or critical feedback, I would no longer be accepted by my peers.	4	/	.93	.94

In order to fully capture students' experience with the imposed transition between anonymous and non-anonymous PA the quantitative data were triangulated with qualitative data from an open-ended questionnaire questions and focus groups. Participants' opinions on the transition from anonymous to non-anonymous were captured via an open-ended question after the first non-anonymous session ("*How did you experience the non-anonymous setting?*") and again after the second non-anonymous session ("*Does your opinion about the importance of anonymity remain or has it changed?*"). Additionally, in the fifth week of the intervention all students were involved in focus groups moderated by the first author. In order to enable rich group discussions the students were split in 4 groups of approximately 11 students. The focus groups were organized around statements. All focus groups were filmed. For the aims of this paper we analyzed results on the statement about the transition approach from an anonymous to a non-anonymous PA setting.

Data analysis

Regarding RQ1 the qualitative content data was treated quantitatively. Repeated measures ANOVAs were performed for all content categories with estimable amounts of feedback messages. The mean number of segments of a specific category was entered as a dependent variable. Furthermore, group (group A or B), gender and presentation mode (whether the student him/herself gave his/her workshop in an anonymous or non-anonymous session) was entered as between subjects variable. For clarity reasons only when significant differences or interaction effects occurred these are discussed in the results section. The category 'general' was not included as it was not identified among our data. Furthermore, as only a small amount of neutral verifications was found and not during each session, these were not presented in the analyses.

Likewise, results of RQ2, partly RQ3a, 3b and 3c were also analyzed via ANOVA repeated measures analyses. The qualitative data from the open-ended question (RQ3a) was brought together and organized in a five column report with 1) the respondent's ID, 2) response to open-ended question 1, 3) response to open-ended question 2, 4) the evolution in stances towards anonymity (i.e. consequently pro or contra anonymity or a change in opinion towards anonymity) and 5) thematic coding of students' arguments for stances towards anonymity. Through a deductive analysis approach the used thematic codes were based on the concepts that were discussed in the theoretical framework: PF quality, perceived PF skills, accuracy and interpersonal variables**.** This approach is particularly useful when one has specific research questions that already identify the main themes or categories used to group the data and then look for similarities and differences (Braun & Clarke, 2006).

The video recordings of the four focus groups were also thematically analyzed. As mentioned earlier, only participants' responses on how they appreciate the transition approach were analyzed. The moderator used a hand raising approach (pro/con) to identify students' opinions, this made it possible not only to get insight on several arguments but also a clear 'final' individual opinion on the discussed topic.

Results

RQ1: What is the evolution in PF quality over time when students practice PA consecutively in an anonymous and non-anonymous setting**?**

First the results about the verifications. In line with our hypothesis (H1) the number of negative verifications of the content-related significantly increased over time [$F(2.36, 99.02) = 3.18$, $p = .038$, $\eta_G^2 = .07$] (Greenhouse–Geisser estimates of sphericity) (see Table 3). This means that after

multiple sessions students gradually dare to indicate more weaknesses in a peers' work regarding the application of didactic principals. More specifically, contrast-analyses revealed a significant increase between Session 1 and Session 2 [$F(1, 42) = 4.94$, $p = .032$, $r = .32$] and Session 1 and 4 [$F(1, 42) = 10.83$, $p = .000$, $r = .45$]. There was no significant difference between Session 2 and 4 [$F(1, 42) = 1.37$, $p = .249$]. In relation to the frequency of positive verifications for the content-related criteria a likewise evolution was found: a significant increase over time [$F(2.49, 104.47) = 7.21$, $p = .00$, $\eta_G^2 = .14$] (Greenhouse–Geisser estimates of sphericity), with significant contrasts between Session 1 to Session 2 [$F(1, 42) = 18.29$, $p = .000$, $r = .56$] and between Session 1 to Session 4 [$F(1, 42) = 16.21$, $p = .000$, $r = .53$]. This means that students' evolution in giving negative and positive verifications for the content-related criteria is comparable.

Similar to the content-related criteria, the negative verifications for the presentation-related criteria augment over time [$F(3, 126) = 6.32$, $p = .000$, $\eta_G^2 = .08$]. Between Session 1 and Session 2 there is a significant increase [$F(1, 42) = 8.78$, $p = .005$, $r = .42$]. It is noted that the results show a significant decrease after the first non-anonymous session [$F(1, 42) = 10.14$, $p = .003$, $r = .44$], but significantly increase between Session 3 and 4 [$F(1, 42) = 9.30$, $p = .004$, $r = .43$], resulting in same quantity as the second anonymous session. Regarding the positive verifications for the presentation-related criteria there was a mean effect of Time [$F(3, 126) = 5.19$, $p = .002$, $\eta_G^2 = .10$], more specifically there was a significant increase between Session 1 and 4 [$F(1, 42) = 5.24$, $p = .027$, $r = .33$], and between Session 2 and 3 [$F(1, 42) = 6.27$, $p = .016$, $r = .36$].

Table 3
Verification type: mean amount of positive and negative verifications per student per session for content- and presentation-related criteria

		Anonymous		Non-Anonymous	
		Session I	Session II	Session III	Session IV
Verification type		M*(SD)*	M*(SD)*	M*(SD)*	M*(SD)*
Content-related criteria	Positive	.82(.66)^{n,o}	2.00(1.61)^{n}	1.75(1.35)	1.64(1.30)^{o}
	Negative	.16(.37)^{p,q}	.39(.54)^{p}	.34(.65)	.57(.85)^{q}
Presentation-related criteria	Positive	3.32(2.90)^{r}	4.32(2.48)^{s}	5.75(3.10)^{s}	5.02(3.46)^{r}
	Negative	.93(.97)^{v,y}	1.66(1.33)^{v,w}	.89(1.06)^{w,x}	1.66(1.16)^{x,y}

Note: same superscripts indicate significant differences at $p < .05$

Regarding elaborations (Table 4), the number of informative elaborations of the content-related criteria significantly increased over time [$F(2.36, 99.02) = 7.67$, $p = .001$, $\eta_G^2 = .14$] (Greenhouse–Geisser estimates of sphericity) (see Table 4). This means that after multiple sessions students gradually give more relevant information to help their peers in error correction (i.e. the application of a certain didactic principle). Additionally, contrast-analyses revealed a

significant increase between Session 1 and Session 2 [$F(1, 42) = 21.95$, $p = .003$, $r = .59$] and Session 1 and 4 [$F(1, 42) = 7.93$, $p = .007$, $r = .40$]. There was a significant decrease between Session 2 and 4 [$F(1, 42) = 4.86$, $p = .033$, $r = .32$].

For the presentation-related criteria, again, students' informative elaborations augment over time [$F(3, 126) = 7.90$, p = .000, $\eta_G^2 = .14$]. More specifically, there is an increase between Session 1 and Session 2 [$F(1, 42) = 18.95$, p = .000, $r = .56$] and between Session 1 and Session 4 [$F(1, 42) = 17.76$, p = .000, $r = .55$].

Concerning suggestive elaborations for the content-related criteria, a significant main effect of time was found [$F(1.32, 55.52) = 15.65$, p = .00, $\eta_G^2 = .26$] (Greenhouse–Geisser estimates of sphericity). However, no suggestions for improvement on the content-related criteria were given during the first session. Student gave significantly more suggestions for improvement on peers' work from Session 3 compared to the anonymous Session 2 [$F(1, 42) = 4.64$, p = .037, $r = .32$], as well in Session 3 compared to Session 4 [$F(1, 42) = 10.35$, p = .002, $r = .45$].

For the presentation-related criteria, main effect Time: [$F(3, 126) = 8.89$, p = .000, $\eta_G^2 = .16$], a significant increase was found between Session 1 and Session 2 [$F(1, 42) = 11.62$, p = .001, $r = .47$], as well between Session 1 and Session 4 [$F(1, 42) = 12.79$, p = .001, $r = .48$].

Table 4
Elaboration type: mean amount of informative and suggestive elaborations per student per session for content- and presentation-related criteria

Elaboration type		Time 1 M*(SD)*	Time 2 M*(SD)*	Time 3 M*(SD)*	Time 4 M*(SD)*
Content-related criteria	Informative	.30(.51)^{p,q}	1.36(1.56)^{q,r}	1.09(1.25)	.75(1.06)^{p,r}
	Suggestive	n.a.	.02(.15)^{s}	.18(.45)^{s,t}	.75(1.06)^{t}
Presentation-related criteria	Informative	.64(1.18)^{u,v}	2.09(2.07)^{u}	1.41(1.62)	1.86(1.82)^{v}
	Suggestive	.14(.55)^{w,x}	.68(.80)^{w}	1.11(1.30)	.68(.93)^{x}

Note: same superscript indicate significant differences at p < .05

In sum, it was expected that negative verifications and informative and suggestive elaborations would increase as it was the case between Session 1 and Session 2 (anonymous phase). The difference in content quality between the anonymous phase (Session 2) and the end of the non-anonymous phase (Session 4) is marginal. Remarkably, there was a decrease in the amount of informative elaborations for the content-related criteria between the second anonymous setting and the fourth non-anonymous session.

RQ2: Do students' perceived PF skills change over time?

When assessing students' PF skills perception, before, during and after the PA-sessions, means (standard deviations in parentheses) were 6.52 (.91), 6.99 (.91), 7.08 (1.00), and 7.11 (.92), respectively. A repeated measures analysis indicates a significant main effect of time [F(2.32, 78.84) = 10.63, p = .000, η_G^2 = .15] (Greenhouse–Geisser estimates of sphericity). Contrast analyses reveal that students perceived a significant improvement of their feedback skills from Session 1 to Session 2 [F(1,34) = 17.03, p = .000, r = .58], but after Session 2 no further significant increases were reported [F(1,34) = .98, p = .329].

RQ3a: How does the transition from an anonymous to a non-anonymous PA affect students' perceptions regarding the attributed importance of anonymity?

Regarding students' importance level towards anonymity (Mean effect Time: [F(2,78) = 6.98, p = .002, η_G^2 = .12]), the pre-test results (Time 1; 4.95 (1.31)) indicate that students initially highly prefer an anonymous PA environment. There was no significant increase after the anonymous sessions (Time 2; 5.25 (1.31)), and there was a significant decrease after the non-anonymous sessions compared to both Time 1 [F(1,39) = 4.74, p = .036, r = .33], and Time 3 4.33 (1.47) [F(1,39) = 17.75, p = .000, r = .56]. One could state that students' importance level evolves towards a more neutral stance.

As the quantitative results show that all students highly prefer anonymity in the pre-test and there is a significant decrease in students' attributed importance towards anonymity towards a more neutral stance, students' responses to the qualitative data will allow to get a detailed picture of students' arguments for this reported decrease. When evaluating students' responses to both the open-ended question on this issue, which was asked after the first non-anonymous session and the second non-anonymous sessions, 4 different kinds of experienced evolutions were found. A first group of students (N = 23) prefer a continued anonymous PA setting both after the first and second non-anonymous setting. The most important reasons that were given, if not in combination, are the fact that a) they feel more comfortable doing it anonymous (N = 10) b) they experienced to be more hesitant to speak freely in the non-anonymous session but that the content of their PF messages was the same. As they can be more straightforward in an anonymous setting they prefer to keep it anonymous (N = 9). A third reason that was stated frequently was the amount of fear for negative consequences in a non-anonymous setting (N = 9). A smaller group of students also indicated that they experienced that they were more honest in an anonymous setting, therefore they would maintain it. A second group of students (N = 10) indicated after the first non-anonymous session that anonymity was not that important anymore, and confirmed this

in their responses after the second non-anonymous session. The majority (*N* = 6) states that the content of their feedback messages is the same, but they spend a bit more time on word choice and nuance. For that reason, students state that for future likewise PF settings non-anonymous participation would not be seen as a burden. Four students mentioned that they came to the understanding after the non-anonymous sessions that it is important to give non-anonymous honest feedback because it otherwise loses its relevance. A third small group experienced (*N* = 9) stated after the first non-anonymous session that anonymity was important, but their opinion changed after the last session. These students thus experienced an evolution throughout the sessions. The reasons for this change were that students stated over time they felt they were giving better argued feedback in the non-anonymous session (*N* = 5), felt more comfortable giving non-anonymous feedback (*N* = 2) and through the experience of the second non-anonymous session experienced an increase of trust in one's own evaluative capabilities (*N* = 2). One student attributed less importance towards anonymity after the first non-anonymous session, but her importance level increase again after the second non-anonymous session. Main raison was that she felt uncomfortable when she was not able to formulate a suggestion for improvement.

Regarding the results of the focus groups, focusing more on the experience as a whole, students' opinions about the transition from an anonymous to a non-anonymous PA reveal that only three of the forty-six students did not agree that this helps to gradually evolve towards the aimed-for interactive PF setting, knowing that they will experience it in real-world (work)contexts. Students' most important motives supporting the approach were related to the effects of negative interpersonal variables (e.g., friendship-marking), which were less present than initially expected by the students (see also RQ3b). Another reason was that the anonymous sessions gave them time to practice their critical FB skills, which they recognized is a skill you need to learn. Finally, students mentioned that after experienced both anonymous and non-anonymous sessions the non-anonymous input was a good starting point to guide the guided teacher discussion part. The following statements illustrate students' opinion on the transition:

Pro transition:

"It is a good thing that we learn how to cope with giving and receiving feedback. The transition makes you consciousness of the fact that you need to learn how to be specific in your FB and there is no need to put a gloss on it, otherwise it becomes useless" (ID 07)

"[…] [the anonymous sessions] were good because we didn't know our peers that well in the beginning and in that way we could still give honest FB. This also allowed me to formulate critical FB. The third session our FB was non-anonymous, which was bit of shock in the beginning, but in the end everyone gave ones' honest opinion on each other's workshop" (ID 31)

Against transition:

"I think the transition is disadvantageous. Not in a way that I gave a different kind of FB, but due to this I really hoped that I would not hurt my peers with my FB, to the point that they would no longer like me " (ID01)

RQ3b: How does the transition from an anonymous to a non-anonymous PA affect students' perceptions regarding their perceptions towards interpersonal variables?

Students' perceptions towards *psychological safety* indicate that students felt initially already moderately comfortable to give their opinion on a peers' work. Overall no significant increase over time was found [$F(2,80) = 1.72$, p= .186], and surprisingly students who gave their workshop in the non-anonymous setting reported an overall significantly higher level of psychological safety [$F(1,40) = 8.40$, p = .006, $r = .42$] (mean difference: .54).

Table 5
Descriptives interpersonal variables

	Time 1	Time 2	Time 3
Trust in Own Evaluative Capabilities	5.02(.77)	5.28(.57)	5.28(.52)
Trust in Peers' Evaluative Capabilities	5.12(.79)	5.29(.79)	5.21(.67)
Psychological Safety	4.61(.95)	4.91(.86)	4.77(.89)
Value Congruency	4.63(.74)	5.23(.67)	5.16(.69)
Friendship marking	/	2.22(.81)	2.82(1.17)
Fear of Disapproval	/	3.60(1.28)	3.24(1.38)

Regarding students' *trust in their own evaluative capabilities* (main effect time [$F(1.70,66.22) = 13.30$,p= .000, $\eta_G^2 = .06$] - Greenhouse–Geisser estimates of sphericity), a significant increase was found between Time 1 and 2 [$F(1,39) = 4.91$, p = .033, $r = .33$], although the initial trust was already high. No further change was noted between the anonymous and non-anonymous settings. Regarding students' *trust in peers' evaluative capabilities* no significant changes over time were found [$F(2,78) = .864$, p= .426].

Regarding students' *value congruency about the PA criteria* a positive evolution over time was found [$F(2,78) = 14.13$, p= .000, $\eta_G^2 = .20$]. More specifically, a significant increase was found from Time 1 to Time 2 [$F(1,39) = 22.19$, p = .000, $r = .60$].

Although the level of perceived *friendship marking* was low after the non-anonymous sessions, contrary to our hypothesis, a significant increase was found between the anonymous and non-anonymous sessions [$t(41) = 2.24$, p =.031, Cohen's $d = .34$].

As expected, students' level of *fear of disapproval* significantly diminished after the non-anonymous (Time 3) sessions when compared to the preceding anonymous sessions (Time 2) [$t(43) = -3.734$, p =.001, Cohen's $d = .56$].

RQ3c: How does the transition from an anonymous to a non-anonymous PA affect students' perceptions regarding their general conceptions towards PA?

Concerning students' conceptions towards PA a significant main effect of time was found [$F(2,70) = 13.30$, p= .000, $\eta_G^2 = .17$]. Means (with standard deviations in parentheses) were 4.90 (1.02), 5.39 (.80), and 5.53 (.71), respectively. As expected a significant increase was found between Time 1 and 2 [$F(1,35) = 13.57$, p = .001, $r = .53$] and Time 1 and 3 [$F(1,35) = 21.29$, p = .00, $r = .61$]. Furthermore, the expected increase between Time 2 and 3 was non-significant [$F(1,35) = 1.44$, p = .238]. This means that during our intervention with two phases there was no decrease in students' conceptions towards PA.

Discussion

The purpose of this study was to explore the effects of disappearing anonymity in PF quality while exploring students' conceptions about the process. Regarding our first hypothesis (H1) about the evolution of PF quality over time, the results regarding verifications and elaborations will be discussed separately. Regarding the verifications it was found that the PF content quality increased as students offered more positive and negative verifications over time which gives support to our implementation of fading anonymity. Interestingly the amount of negative verifications of the presentation-related criteria significantly decreased in the session after the transition from anonymous to non-anonymous sessions. This finding suggests that students might benefit of experiencing both type of settings in order to non-anonymously point negative aspects of their peers' work. Furthermore, the amount of negative verifications in the second non-anonymous session becomes equal to or higher than the negative verifications in the second anonymous session. This finding again favors our transition approach because the feedback becomes more descriptive of the actual performance rather than only pointing out to the positive aspects. The fact that amount of positive verifications remains stable (content-related criteria) or increases (presentation-related criteria) during the non-anonymous sessions points at the fact that students feel the need to pinpoint both positive and negative aspects of a peers' work, which corroborates findings by Gielen and De Wever (2015).

Regarding the elaborations, our hypothesis of an overall increase was also confirmed. The significant decrease of informative elaborations between the second anonymous session and the second non-anonymous session might be related to the fact that due to multiple experiences with the workshops, students gave significantly more elaborative suggestions to improve a peers' work, rather than informing them why certain aspects of their work were positive of negative. This points at an improvement of students' evaluative expertise over time (Sadler, 2010) as they relate their evaluations to the evaluations of others: they reflect whether their judgements were appropriate or not, looking for ways to improve future feedback content and wonder what they have missed in making their judgements that others have noticed (Boud, Lawson, & Thompson, 2013).

Our second hypothesis (H2) about students' perceived evolution in PF skills can be partially maintained. As students' already rate themselves highly positive in Session 1 there was only a significant increase towards Session 2. More importantly, there is no decrease in perceived improvement after Sessions 3 and 4. This suggests that students do not feel hindered by the non-anonymous setting that was created in Session 3 and 4.

As expected the questionnaire data show that students' importance level significantly decreased after Session 2 (anonymous) (H3a). This means that overall students' opinion on the importance of anonymity resulted in a rather neutral stance. The qualitative data however showed a more diverse image: half of the students preferred a continued anonymous PA setting both after the first and the second non-anonymous setting. The second half of the studied population found anonymity less important after the first or second non-anonymous session. The data of the focus groups clearly shows that students differentiate between the amount of importance they attribute towards anonymity and whether a transition approach from an anonymous to a non-anonymous is seen as a good approach to evolve towards direct interactive non-anonymous feedback settings. Only three out of the forty-six students did not agree on this. Although our initial hypothesis of an overall increase of students' PA conceptions could not be confirmed, the fact that students' PA conceptions remained highly positive after the non-anonymous phase was initiated, again favors the implementation of a transition approach.

Although previous research pointed out that anonymity might help to reduce interpersonal burdens, the results of this study could not confirm nor reject these findings (RQ3b). That is, students' perceptions towards positive interpersonal variables (psychological safety, trust in own and others' evaluative capabilities and value congruence) were initially already – moderately – positive. Only for trust in own' evaluative capabilities and value congruency an increase was found after the anonymous sessions. Yet again, our transition approach to a non-anonymous setting had no negative effect on the positive interpersonal variables. Regarding the negative interpersonal

variables, the opinions regarding friendship marking and fear of disapproval were low after the anonymous session and remained low after the non-anonymous sessions. These findings might be explained by implemented PA scaffolds (i.e. active involvement in rubric criteria development and guiding questions) and the fact that our sample already had already positive attitudes towards PA (as also confirmed by the pre-test results on PA conceptions) and these positive attitudes helped to overcome interpersonal burdens in all phases of the intervention. The fact that almost all students highly appreciate our transition approach strongly suggests that it meets their need for practice in a safe environment. Moreover, as students did not really expect an influence of interpersonal variables, it can be expected that in settings in which this interpersonal burden is more present the application of this transition approach might be even more valuable.

Implications

Our findings are important for educational practice as they add to our highly needed understanding of PA as a powerful pedagogical practice (Panadero & Brown, 2016). This study confirms earlier findings that practice is an important component in PA implementations (Gielen & De Wever, 2015; Liu & Carless, 2006). Second, when a transition from an anonymous to a non-anonymous PA environment is facilitated, students' PF quality in the anonymous phase increases over time and the PF quality in the non-anonymous sessions eventually becomes comparable. As recently stated by Panadero and Brown (2016) it is important that teachers themselves practice PA with other teachers for their professional development in order to give them greater awareness of the interpersonal dynamics within PA. Exploring our transition approach with anonymity in different contexts and with different groups is certainly part of that. This would help teacher to decide whether they will opt for anonymous modes of feedback depending on the time available and the specific characteristics of the student population.

Limitations and future research

Given the sample size, gender bias of the sample (mostly female) and the fact that this was a peer group assessment setting, the findings of this research should be interpreted with caution. In this study PF quality was defined based on the presence of structural components in a PF message as defined by Gielen and De Wever (2015). As no impact of the received PF on likewise future performances was taken into account (Evans, 2013), the current quality measurement can only be interpreted in terms of its potential impact for future performances. Additionally, the interactive exchange during the oral discussion with the teacher was not actively monitored. Future research could analyze the interactive exchange between peers during the guided teacher

discussions as it will allow to get a deeper understanding of how dialogic feedback arises (Carless, 2010).

Conclusion

The creation of a safe and supportive learning environment in which students feel comfortable and confident to assess their peers, is essential for the quality of PA activities. Although offering anonymity to the assessors within these activities has been recommended in previous research for its positive effect on students' PA conceptions and interpersonal burdens (e.g. van Gennip, Segers, & Tillema, 2009), currently research on the actual feedback behavior of students in anonymous PA settings was lacking. Moreover, the use of anonymity as a temporary scaffold to gradually evolve towards a dialogic feedback environment (e.g. Howard et al., 2010) had not yet been explored.

The content analysis of the PF messages revealed that the quality of the PF increases in the anonymous phase and that over time the FB quality in the non-anonymous sessions was comparable. Focus group results confirm that students appreciate the anonymous phase to practice their PF skills, in order to produce high quality feedback. Our findings suggest that anonymity can be used as a valuable scaffold to ease students' importance level towards anonymity and their associated need for practice in a safe environment. Consequently, this study in a synchronous setting points out that depending on the available time to organize PA and the specific characteristics of the student population (i.e. in groups in which interpersonal burdens might hinder the PA activity), teachers could choose between only anonymous PA (short) or sequence of anonymous and non-anonymous PA (long) depending whether they intent to focus on the creation of a safe learning environment, or work towards a dialogic non-anonymous PA environment.

References

Ainsworth, S., Gelmini-Hornsby, G., Threapleton, K., Crook, C., O'Malley, C., & Buda, M. (2011). Anonymity in classroom voting and debating. *Learning and Instruction*, *21*, 365–378. doi:10.1016/j.learninstruc.2010.05.001

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing Procedures for Implementing Peer Assessment in Large Classes Using an Action Research Process. *Assessment & Evaluation in Higher Education*, *27*(5), 427–441. doi:10.1080/0260293022000009302

Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151–167. doi:10.1080/713695728

Boud, D., Lawson, R., & Thompson, D. G. (2013). Assessment & Evaluation in Higher Education Does student engagement in self- assessment calibrate their judgement over time ? *Assessment & Evaluation in Higher Education*, 37–41. doi:10.1080/02602938.2013.769198

Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 1–14. doi:10.1080/02602938.2015.1018133

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(May 2015), 77–101. doi:10.1191/1478088706qp063oa

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*. doi:10.1080/03075071003642449

Cho, Y. H., & Cho, K. (2010). Peer reviewers learn from giving comments. *Instructional Science*, *39*(5), 629–643. doi:10.1007/s11251-010-9146-1

Falchikov, N. (1995). Peer Feedback Marking: Developing Peer Assessment. *Innovations in Education & Training International*, *32*(2), 175–187. doi:10.1080/1355800950320212

Fastré, G. M. J., van der Klink, M. R., Sluijsmans, D., & van Merriënboer, J. J. G. (2013). Towards an integrated model for developing sustainable assessment skills. *Assessment & Evaluation in Higher Education*, *38*(5), 611–630. doi:10.1080/02602938.2012.674484

Freeman, M., & McKenzie, J. (2000). Self and Peer Assessment of Student Teamwork: Designing, implementing and evaluating SPARK, a confidential, web based system. In *Flexible Learning for a Flexible Society*. Retrieved from http://ascilite.org.au/aset-archives/confs/aset-herdsa2000/procs/freeman.html

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, *52*, 315–325. doi:10.1016/j.chb.2015.06.019

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, *36*, 101–111. doi:10.1016/j.tate.2013.07.008

Hattie, J., & Gan, M. (2011). Instruction Based on Feedback. In *Handbook of Research on Learning and Instruction* (pp. 249–270). Routledge. doi:10.4324/9780203839089.ch13

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. doi:10.3102/003465430298487

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, *38*, 21–27. doi:10.1016/j.stueduc.2012.04.001

Howard, C. D., Barrett, A. F., & Frick, T. W. (2010). Anonymity to Promote Peer Feedback: Pre-Service Teachers' Comments in Asynchronous Computer-Mediated Communication. *Journal of Educational Computing Research*, *43*(1), 89–112. doi:10.2190/EC.43.1.f

King, A. (2002). Structuring Peer Interaction to Promote High-Level Cognitive Processing. *Theory Into Practice*, *41*(1), 33–39. doi:10.1207/s15430421tip4101_6

Li, L. (2016). The role of anonymity in peer assessment. *Assessment & Evaluation in Higher Education*, *2938*(April), 1–12. doi:10.1080/02602938.2016.1174766

Liu, N.-F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, *11*, 279–290.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102–122. doi:10.1080/02602938.2013.795518

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Human factors and social conditions of assessment. New York: Routledge* (pp. 1–39). New York, NY: Routledge.

Panadero, E., & Brown, G. T. L. (2016). Teachers' reasons for using peer assessment: positive experience predicts use. *European Journal of Psychology of Education*. doi:10.1007/s10212-015-0282-5

Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, *39*(4), 195–203. doi:10.1016/j.stueduc.2013.10.005

Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. (2010). Formative peer assessment in a CSCL environment: a case study. *Assessment & Evaluation in Higher Education*, *30*(4), 417–444. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/02602930500099219

Raes, A., Vanderhoven, E., & Schellens, T. (2013). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Studies in Higher Education*, *0*(0), 1–16. doi:10.1080/03075079.2013.823930

Reinholz, D. (2015). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 1–15. doi:10.1080/02602938.2015.1008982

Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2017). "Now you know what you're doing right and wrong!" Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education*, 1–21. http://doi.org/10.1007/s10212-017-0329-x

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, *35*, 535–550.

Sluijsmans, D. M. A. (2002). Student involvment in assessment: The training of peer assessment skills. (Unpublished doctoral dissertation). Open University of the Netherlands, Heerlen.

Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: are they critical for feedback perceptions and efficiency? *Learning and Instruction*, *20*, 291–303.

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, *20*, 265–269.

Sung, Y.-T., Chang, K.-E., Chang, T.-H., & Yu, W.-C. (2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence*, *33*, 135–145. doi:10.1016/j.adolescence.2009.04.004

Topping, K. (2009). Peer Assessment. *Theory Into Practice*, *48*(1), 20–27. doi:10.1080/00405840802577569

Topping, K. J. (1998). Peer Assessment between Students in Colleges and Universities. *Review of Educational Research*, *68*(3), 249. doi:10.2307/1170598

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*(4), 339–343. doi:10.1016/j.learninstruc.2009.08.003

Tsai, C.-C., Lin, S. S. ., & Yuan, S.-M. (2002). Developing science activities through a networked peer assessment system. *Computers & Education*, *38*(1-3), 241–252. doi:10.1016/S0360-1315(01)00069-0

Van der Pol, J., Van den Berg, B. A. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education*, *51*, 1804–1817.

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, *4*, 41–54. doi:10.1016/j.edurev.2008.11.002

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, *20*, 280–290. doi:10.1016/j.learninstruc.2009.08.010

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, *81*, 123–132. doi:10.1016/j.compedu.2014.10.001

Yu, F.-Y., & Liu, Y.-H. (2009). Creating a psychologically safe online space for a student-generated questions learning activity via different identity revelation modes. *British Journal of Educational Technology*, *40*(6), 1109–1123. doi:10.1111/j.1467-8535.2008.00905.x

Yu, F.-Y., & Sung, S. (2015). A mixed methods approach to the assessor's targeting behavior during online peer assessment: effects of anonymity and underlying reasons. *Interactive Learning Environments*, 1–18. doi:10.1080/10494820.2015.1041405

Appendix 1

Rubric criteria

Presentation-related criteria
14. Body language
15. Use of voice and language
16. Interaction with audience
Content-related criteria:
17. Principle of illustration and specificity
18. Principle of activity
19. Principle of gradualism

Example: Principle of gradualism

Score 1	Score 2	Score 3	Score 4	Score 5
There is a total lack of structure, which makes it impossible for the audience to understand the content. The group lacked time management.	There was an attempt to structure the workshop, but they failed. Time management was poor.	The material was logically built up, though its presentation lacked structure. The session ended on time, though the use of time wasn't efficient.	There was a good build-up and structure. The group presented the material in an effective way. Time management was good.	The workshop had a clearly structured build-up. Time management was optimal.

7 General discussion and conclusion

Chapter 7
General discussion and conclusion

Introduction

It is well recognized that assessment is a crucial driver of student learning and that well-implemented assessment processes provide prospects for meaningful learning (e.g., Black & Wiliam, 1998; Carless, 2017). Developments in the field predominantly stress on the importance of formative approaches, where the information that has been gathered through the assessment tasks provides feedback to the students to steer their learning process. Parallel to and building upon the theory of formative assessment, a number of authors (e.g., Boud & Soler, 2015; Fastré, van der Klink, Sluijsmans, & van Merriënboer, 2013) have taken up the idea of sustainable assessment. For these authors, formative assessment is promising in theory but, on its own, is insufficient in helping students to learn to assess their own learning and does not prepare them to act as lifelong learners beyond graduation and throughout their career. In a complex world with rapidly changing professions, people need to be able to recognize when further improvement in one's own performance or the learning of new skills is required (Regehr & Eva, 2006). This approach takes a practical view that considers assessment as a key element of the process of developing and informing the learner's judgment for learning beyond the immediate task (Boud, 2007). These assessment skills help students become self-regulated learners who are aware of their own qualities and shortcomings and who know how to overcome potential hurdles (Butler & Winne, 1995).

Because of the active involvement of students in the assessment process, peer assessment (PA) has been embraced as an innovative method of formative and sustainable assessment and is often seen as a way to offer significant educational value to learning (e.g., Topping, 2010). PA is defined as any educational arrangement in which students judge their peers' performance by providing grades and/or offering written or oral feedback (Topping, 1998). PA has many benefits (Dochy, Segers, & Buehl, 1999; Panadero & Brown, 2017; Topping, 2003) and can be considered a learning tool owing to the active involvement of the learner in the learning process and the fact that it provides the learner with skills to assess criteria that define high-quality work (Orsmond, Merry, & Callaghan, 2004; Topping, 1998). Nonetheless, PA is a difficult and intensive practice to implement. For example, it involves multiple social and human factors that need to be taken into account because peer assessment does not happen in isolation; it produces thoughts, actions, and emotions as a consequence of the interaction of assessees and assessors, which can have an impact on the quality of the PA process (Panadero, 2016), especially in face-to-face classroom contexts (Latané, 1981; Pope, 2005). Additionally, in the pursuit of PA environments in which students feel

comfortable to participate, it has often been stated that offering anonymity to assessors is desirable (Ballantyne, Hughes, & Mylonas, 2002; Vickerman, 2009) or should at least be studied (Howard, Barrett, & Frick, 2010; Topping, 1998). Furthermore, to formulate and articulate judgments on a peer's work, students need to be able to offer high-quality feedback and, therefore, need to develop assessor skills. As such, previous research has indicated that students require practice, training (Sluijsmans, 2002), and guidance (Hovardas, Tsivitanidou, & Zacharia, 2014; Ernesto Panadero & Romero, 2014). This dissertation focuses on the human and social conditions that can stimulate and/or impede the implementation of valuable PA practices by exploring students and teachers' perceptions of possible influencing variables. This is necessary because how this necessary environment can be created is seldom explained in-depth alongside the peer assessment templates that teachers are currently are encouraged to use (Hattie, 2016). This dissertation primarily focuses on perceptions because peer assessment is a social experience – it does not happen in isolation, given that students define and practice assessment from their own point of view and in relation to others (e.g., teachers and peers) (McInerney, Brown, & Liem, 2009).

Based on an adapted version of the conceptual framework, called as the feedback triangle by Yang and Carless (2013), three adapted dimensions for facilitating and studying dialogic PA processes were identified (see Fig. 1):

(a) The first dimension was the *social and interpersonal nature* of PA, indicating the social-affective dimension in which we focus on the human and social conditions that can stimulate and/or impede the implementation of valuable PA practices with the PF component in synchronous PA settings in which immediate peer feedback (PF) is given by means of Mobile Response Technology (MRT). More specifically, we investigated students' perceptions of interpersonal variables (e.g., trust, fear of disapproval, value congruency), anonymity, and accuracy and their relationship with students' beliefs on the educational value of PA (Chapter 2, 4 & 6). Furthermore, Chapter 3 explored teachers' awareness of students' concerns regarding these issues.

(b) Dimension two of the PF triangle was *PF Quality*, indicating the cognitive dimension, which focused on the effect of PA practice on students' PF quality based on the occurrence of structural components within their PF messages (Chapters 5 & 6).

(c) Adapted dimension three was *organization and management* of PA practices, implying the structural dimension, which refers to the organization, timing, sequencing, and modes of feedback, allied to non-disciplinary resources for generating and providing feedback. In this dissertation, we focused on three components that are frequently focused on in the organization and management of these practices: 1) PA task design, 2) enabling technology, and 3) structure and support.

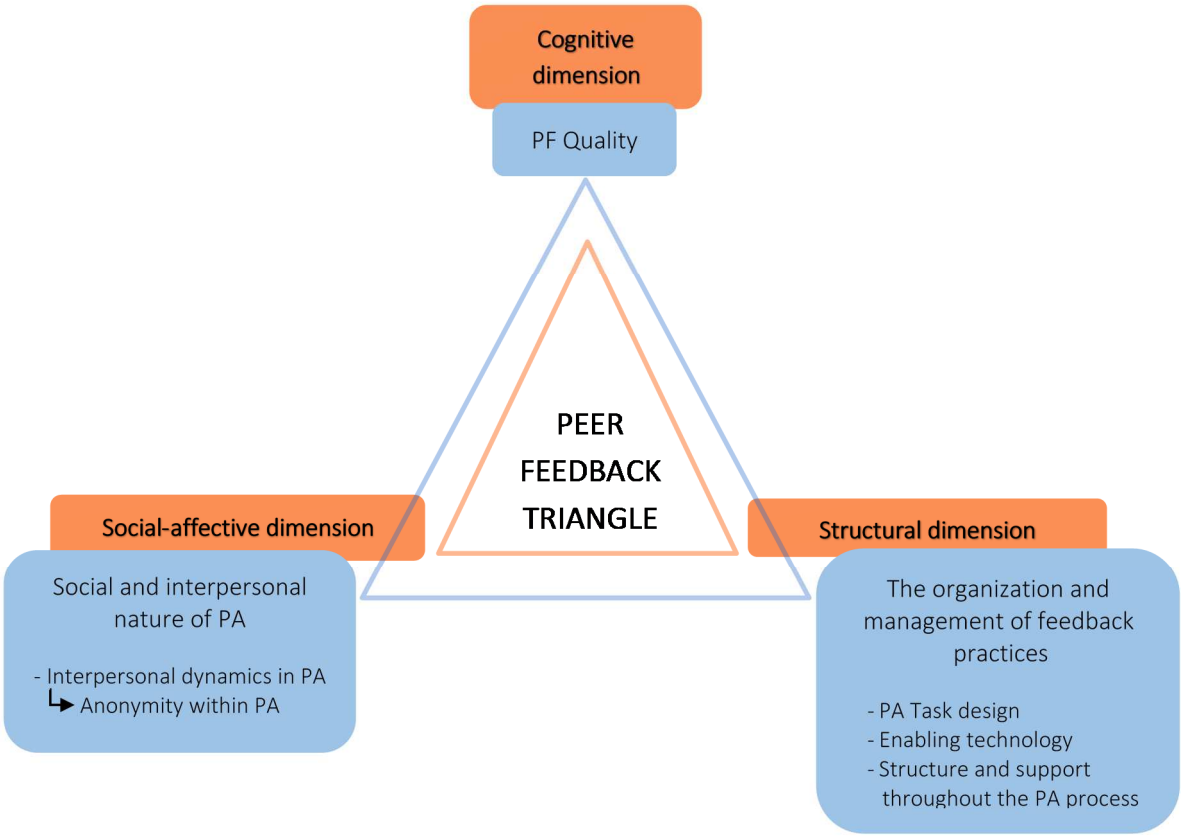


Figure 2. Peer feedback triangle

In the introductory chapter of this dissertation (Chapter 1), research challenges within these dimensions were discussed and the following research objectives were formulated:

Research objective 1 (RO1):	To determine the current use of PA practices in secondary education in Flanders and to explore the relationship between participants' PA conceptions and its social nature	*Social-affective + structural dimension*
Research objective 2 (RO2)*:*	To investigate students' perceptions of interpersonal variables and attributed importance to anonymity when involved in a face-to-face PA facilitated by Mobile Response Technology	*Social-affective + structural dimension*
Research objective 3 (RO3):	To examine students' PF quality within these face-to-face PA settings	*Cognitive dimension*

Results & discussion

In this concluding chapter, an overview and discussion of the main results of the different studies is presented and framed within the three dimensions of the PF triangle.

Social-affective dimension: Social and interpersonal nature of PA

Overview and discussion of the results

In the social-affective dimension of the PF triangle, we focused on the human and social conditions that can stimulate and/or impede the implementation of valuable PA practices in light of the PF component in synchronous settings. More specifically, considering the inherent social and interpersonal nature of PA activities, in two survey studies, we explored students' perceptions of interpersonal processes (Chapter 2) and teachers' awareness of students' perceptions of these processes (Chapter 3) within PA in relation to their beliefs on the educational value of PA (RO1). Furthermore, in two intervention studies within authentic classroom settings in higher education (Chapter 4 & 6), students' perceptions of these interpersonal processes throughout and after their participation in multiple PA sessions within synchronous settings were studied (RO2). Overall, both the large-sample survey studies, as intervention studies, have confirmed the complex nature of PA processes, which can trigger powerful feelings and have an impact on participants' beliefs regarding the educational value of PA. These studies add to our understanding of the impact of students' perceptions of interpersonal variables, accuracy, and anonymity on the perceived educational value of PA.

In **Chapter 2**, the relationship between students' perceptions of interpersonal processes within PA and students' attributed educational value of PA was explored in a survey study involving 3066 secondary education students with previous PA experience. Overall, the majority of the students in all levels had experienced a PA activity at least once. The frequency of experience with PA was generally high, with "more than thrice" being the most frequently chosen option. Students in Level 2 and 3 indicated having experienced predominantly summative, grade-oriented PA. Currently, PA is most frequently applied in the Dutch, French, and math courses. The majority of the students indicated that PA activities were conducted in a non-anonymous mode. The SEM results indicated that four variables significantly predict the educational value students attribute to PA: (1) *trust in evaluative capabilities*, which measures the extent to which students believe in their own and their peers' evaluative capabilities, (2) *negative interpersonal process*, which refers to the extent to which students deem that friendship-based marking is prevalent and

how much they fear disapproval if they give low scores and/or critical feedback, (3) *accuracy*, which refers to students' perceived accuracy of PA and whether they think their peers are capable of giving accurate judgments, and (4) *importance of anonymity within PA*, which indicates the amount of importance attributed to guaranteeing anonymity to assessors, assessees, and/or teachers within PA activities. In comparison with Chapter 2, only one predictor (perceived accuracy) could be identified in **Chapter 3**, where we asked 225 teachers about their awareness of interpersonal processes. The first factor was *interpersonal factors within PA*, which refers to the extent to which teachers have considered the effect of interpersonal processes on the PA process. The results show that teachers in the studied population had moderate awareness of interpersonal processes in PA. The second factor was *importance of anonymity*, which refers to teachers' awareness about students' attributed importance to guaranteeing anonymity of assessors and/or assessees. Similar to the first identified factor, there was moderate awareness of this issue. The third factor was *accuracy*, which measures teachers' perceived accuracy of PA and whether they think students are capable of giving accurate judgments. This variable was moderately positively evaluated. The fourth factor was *educational value of PA,* which refers to the extent to which teachers think PA is a valuable assessment method. The educational value of PA was highly positively appraised by the teachers in our sample. To discuss the results of Chapter 2 and 3, Chapter 2 will be used as a reference study in the next sections, when applicable.

(1) In Chapter 2, our model showed that the amount of *trust in your own and peers' evaluative capabilities* positively contributed to students' perceived educational value of PA. In previous research, reciprocal trust in the assessment skills and capability of peers to give judgment on one's work has been proven to be important within the PA process (Panadero, 2016; van Gennip, Segers, & Tillema, 2010), as it may lead to deeper learning approaches (Cheng & Tsai, 2012). The fact that students in Level 3 had a significantly lower amount of trust than those in Level 1 is a matter of concern. A possible explanation for this is that the current PA implementation in the higher levels of secondary education is still predominantly aimed at scoring, instead of pursuing a "deep" approach with much more attention to the peer feedback component, which was also reported by Panadero and Romero (2014). Nevertheless, this finding calls for constant monitoring and enhancement of the classroom atmosphere underpinned by a philosophy that encourages reflection so that critical feedback will be perceived as being constructive instead of judgmental (Harris & Brown, 2013; Havnes, Smith, Dysthe, & Ludvigsen, 2012; Heitink, Van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016).

(2) Regarding *negative interpersonal processes*, the findings suggest that students' awareness of these processes in PA – including their potential undesirable effects – does not result in a negative stance towards the educational value of PA. Again, this calls for trust building and

awareness raising activities regarding potential negative interpersonal processes with students before their actual participation in PA. The higher the students' awareness on this issue the lower the negative influence one can expect as a consequence of students using suitable coping mechanisms. In contrast to the findings of Van Gennip et al. (2010), *value congruency* and *psychological safety* did not explain the students' PA conceptions. A possible reason behind this is the fact that the study of van Gennip et al. (2010) had a quasi-experimental design, where students' perceptions were measured directly after the intervention they conducted, resulting in a stronger effect of these variables. This reasoning was confirmed in Chapter 4 (see further). In comparison, in the survey with teachers (Chapter 3), the extent to which teachers have considered the effect of interpersonal processes (both positive and negative) on the PA process was not identified as a positive predictor of their attributed value of PA. The results show that teachers in the studied population had moderate awareness of students' concerns regarding interpersonal processes in PA. This indicates a possible lack of insight and mastery of these processes from a teachers' point of view. This finding, therefore, suggests that in line with the aforementioned conclusions in the chapter on students' perceptions (Chapter 2), future research should focus on enhancing teachers' consideration of the human and social nature of these practices as well as on how to create a PA learning environment based on trust and respect (Brown & Harris, 2016).

(3) One crucial PA aspect closely related to the aforementioned variables is *accuracy*. Accuracy was proven to be the most important positive predictor in the structural model. The students in Level 2 and 3 reported significantly lower rates of perceived accuracy than those in Level 1. This finding is worrisome as one would expect that more experienced students become more accurate in their assessment and, thus, perceive the PA procedure as more accurate in the higher levels. A possible explanation could be the fact that the PA practices reported in Level 2 and 3 were predominantly, summative, grade-oriented PA. The fact that students in Level 2 and 3 reported a lower value of PA hints at a confirmation of the aforementioned reasoning.

Furthermore, the importance of this variable in relation to students' attributed educational value of PA was confirmed in the survey involving teachers (Chapter 3), where teachers' perceived accuracy proved to be the solely significant predictor.

(4) With regard to the fourth factor, i.e., *importance of anonymity,* a negative relationship with *perceived educational value* was found. Overall, the use of anonymous modes of PA was very low in the studied sample. This significantly negative relationship with a sample of students that experienced predominantly non-anonymous PA confirms the previous theoretical work on PA by Topping (2010), who indicated that privacy is an important structural component of PA. Importantly, the attributed importance of anonymity was lower for students with three or more than three PA experiences. This finding suggests that practice leads to more willingness to

participate in non-anonymous PA settings, which confirms the need for research on PA settings in which a transition from an anonymous to a non-anonymous setting is made, as established in Chapter 6. The descriptive analysis in the teachers' survey (Chapter 3) showed little use of anonymous modes of PA. Furthermore, teachers' perceived importance of students' attributed importance of anonymity within PA was only moderate, which was in contrast with the high level of students' perceived importance of anonymity as reported by students in Chapter 4 and 6.

In order to determine how students react to and interact with the social nature of PA processes, **Chapter 4** explored 39 higher education students' experiences within a *reciprocal* (students assess each other and all students undertake both the role of assessor and assessee), *anonymous* (single blind: anonymity is offered to the assessor during FB provision), and *synchronous* (immediate with the use of MRT) PA setting in *group presentations*. The results showed that in the PA setting that was created, the students felt safe in terms of interpersonal risk-taking and had a high amount of trust in their own and others' evaluative capabilities. Moreover, there was considerable unanimity with regard to the goals and criteria of the PA task. The majority explicitly mentioned that thanks to the guaranteed anonymity, they gave more honest scores and feedback and were less influenced by bonds of friendship. This was also reflected in the quantitative results, thus confirming the theoretical claims that anonymity can help to overcome interpersonal burdens (e.g., Yu & Sung, 2015). The scores on the variable *fear of disapproval when giving low scores or negative feedback* were rather low, and one-fifth of the participants directly ascribed their low score on this variable to the guaranteed anonymity of the assessor, thus indicating diminished fear of repercussions. Regarding the relationship between all interpersonal variables and general conceptions of PA within an anonymous PA setting, firstly, the results showed that when students reported a high degree of fear of disapproval, they tended to have less positive conceptions of PA. Secondly, the stronger the unanimity of goals and criteria of PA the more positive the students' conceptions of PA. This finding is in contrast with the findings of van Gennip et al. (2010), who found that the lower the value congruency on goals the more positively the students viewed PA; their findings were therefore confuted. The findings also confirm Panadero's (2016) reasoning that our intensive PA practice was able to produce shared goals and criteria. Likewise, a higher amount of trust in the evaluative capabilities of peers predicts more positive conceptions of PA. The results of the study on the relationship between interpersonal variables and conceptions (to learning) in PA partially confirmed the findings of Van Gennip et al. (2010), who explored this relationship in a non-anonymous PA setting in vocational secondary education.

The descriptive results in **Chapter 6** confirmed the findings related to students' perceptions of interpersonal processes, as reported in Chapter 4. The students' perceptions of positive interpersonal variables (psychological safety, trust in own and others' evaluative capabilities, and

value congruence) were initially already – moderately – positive. Only for trust in one's own evaluative capabilities and value congruency, an increase was found after the anonymous sessions in Chapter 6, although these differences were practically irrelevant. Regarding the negative interpersonal variables, the importance given to friendship-based marking and fear of disapproval was found to be low after the anonymous session and remained moderately low after the non-anonymous sessions.

Directions for future research

In sum, two recurrent themes which offer food for thought and directions for future research can be identified in this discussion on the social-affective dimension of the PF triangle: (a) our studies on students' perceptions of interpersonal processes within face-to-face PA activities lead us to the conclusion that focusing on monitoring and enhancing a supportive interpersonal climate might be key to overcoming, what Strijbos (2017) recently called, the interpersonal eddies of PA. (b) Teachers' awareness of these interpersonal issues was found to be moderate and future research should focus on how to raise teachers' awareness of this inherent interpersonal nature of PA.

Building a supportive interpersonal climate

The significant influence of several interpersonal processes within PA on the educational value that students attribute to PA, repeatedly points out to the need for building a supportive interpersonal climate within classrooms. Establishing a supportive classroom environment, in which criticism is welcomed, can make PA activities more constructive (Andrade, 2010; Hattie & Timperley, 2007).

Our intervention studies showed that active involvement of students in the development of assessment criteria and offering multiple PA sessions and guidance questions led to an overall positive PA climate in which students felt safe to participate. This supportive environment should motivate students to express their ideas and encourage learning from error (Brown & Harris, 2014). In that sense, involvement in PA has the potential to promote prosocial behaviors, calming emotions, and a sense of belonging (Rowe, 2017). Nevertheless, expecting students to be automatically immune to negative interpersonal processes, criticism, or any form of negative interaction indicates naïveté and wishful thinking (Brown & Harris, 2016). This focus on the creation of a supportive climate is in line with the recent claims in literature according to which embedding (peer) feedback within learning design and emphasizing interactions between students and teachers (outlined in the *structural dimension* in this dissertation) is essential in the realization of sustainable learning environments (Boud & Molloy, 2013). The emphasis on embedding PA within the learning design implies that PA should no longer be seen as an episodic

mechanism (Panadero, 2016; Rowe, 2017). As a consequence, future research that focuses on supporting students to cope with the inevitable interpersonal effects of participating in PA should look into intervening at classroom level, rather than at the individual level. In other words, building on the work of Pryor and Crossouard (2008), future research should focus on the "meta-social" element in classroom interactions in order to support students' awareness of social aspects related to (peer) assessment (Cowie & Harrison, 2016). As such, the findings in this dissertation on the interpersonal nature of PA, thus, support the recent recommendation in "Assessment for Learning" (AfL) literature to position all classroom interactions, such as PA practices, as teaching-learning or pedagogical interactions (Brown & Harris, 2016). Brown and Harris (2016) correctly argue that: "This approach, while seemingly downgrading classroom interactions from the 'lofty' status of being assessment, would potentially free up such interactions from the burdens associated with evaluation and accountability and focus them more sharply on diagnosing needs and strengths and providing feedback and guidance for improved instruction and learning" (p. 514).

Raising teachers' awareness on the interpersonal nature of PA

As mentioned in Chapter 3, the teachers' stance and actions are critical in developing a classroom culture that supports the enactment of two-way, dialogic PF practices. As a first important step in examining this complex issue, teachers' awareness of students' concerns regarding interpersonal processes in PA was studied. However, the results showed a lack of teachers' awareness of the students' concerns, and this lack might lead to an aggravation of affective threats such as discomfort, fear of disapproval, and/or distrust (Higgins, Hartley, & Skelton, 2001). To prevent this, teachers will have to show interpersonal caring (e.g., actively build up trust) (see structural dimension) and also cultivate suitable conceptions of assessment (i.e., assessment is for enhancing students' learning) (Brown, 2011). Boud (2016) states that the development of suitable conceptions precedes specific assessment practices and, as a consequence, is the key element of pedagogical practices. Therefore, studying teachers' conceptions of assessment is important at a time when innovation of assessment practices is on the educational agenda. This specific focus can be a starting point for the development of professionalization programs that expressly aim to make teachers aware of their conceptions as well as changing those conceptions that do not match the aimed-for innovation (Segers & Tillema, 2011). To date, the conjunction of assessment literacy (i.e., understanding and use of assessment and data from assessment) and conceptions of assessment has not been adequately explored. We think this is a valuable direction to explore because conceptions of assessment denote the belief systems that teachers have about the nature and purposes of assessment and also encompass their cognitive and affective responses (Xu & Brown, 2016). Teachers tend to adopt new knowledge, ideas, and strategies of assessment that

are congruent with their conceptions of assessment, while rejecting those that are not (Hill, 2011; Xu & Brown, 2016). In other words, assessment conceptions can hinder the operationalization of adequate assessment methods. Previous research has proven the robustness of these conceptions which may have to do with each participant's personal history and personal relationship with assessment derived from their own assessment career (Ecclestone & Pryor, 2003). Professionalization initiatives on assessment must, thus, address preexisting conceptions and their causes (Brookhart, 2011). In doing this, we must avoid that *assessment literacy* is enhanced through professionalization, but *conceptions of assessment* that may influence the practical application of (peer) assessment remain (Deneen & Brown, 2016).

Cognitive dimension: The quality of PF

Overview and discussion of results

In the cognitive dimension of the PF triangle, we examined the effect of PA practice on students' PF quality based on the occurrence of structural components (i.e., verifications and elaborations) within PF messages (RO3) (Hattie & Gan, 2011; Narciss, 2008; Shute, 2008). As such, the studies in Chapter 5 and 6 are the first that explored students' development of evaluative expertise within face-to-face synchronous PA settings in which offering anonymity to the assessors was examined. The main results will be presented in this section. Additionally, it was examined whether students perceived an improvement in their PF skills as compared to the actual evolution of their PF quality. Furthermore, as students' response to PF is not only a feature of the activities themselves but also depends on the perceived usefulness of PF, this self-report measure was included in Chapter 5.

Previous research claims that successful feedback messages should contain both verifications and elaborations (e.g., Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Gielen & De Wever, 2015). With regard to verifications, previous studies have indicated that students usually tend to give mostly positive verifications on a peer's performance (i.e., making a positive evaluative statement on past performance) (Gielen & De Wever, 2015). Looking at the results of the content analyses that were performed in this dissertation, with regard to positive verifications, it can be concluded that the number of positive verifications remained stable throughout multiple PA sessions (in a secondary education PA setting, as seen in Chapter 5) or increased (in a higher education setting, as seen in Chapter 6), thus confirming the findings of Gielen and De Wever (2015). This is not problematic as long as the other PF components are also present, as will be discussed below. In line with our hypothesis in both studies, PA practice leads to an overall – i.e., from start to end – increase in negative verifications (i.e., making a negative evaluative statement). Interestingly, in Chapter 6, after the transition from anonymous to non-anonymous sessions (session 3), there was a decrease in the amount of negative verifications. Moreover, the amount of negative verifications in the second non-anonymous session (session 2) was equal to or higher than the negative verifications in the preceding second anonymous session (session 4). This progression suggests that students benefit from experiencing anonymous and non-anonymous settings successively, as they are, subsequently, able to point out negative aspects of their peers' work, even in non-anonymous settings. This finding supports the transition approach.

With regard to the evolution in the elaborations within PF messages, in Chapter 5 and 6, an overall significant increase in both informative and suggestive elaborations was confirmed. More specifically, in Chapter 6, the significant decrease in informative elaborations from the second anonymous session to the second non-anonymous session might be related to the fact that owing to multiple experiences with the workshops, the students gave more elaborative suggestions to improve a peer's work rather than informing them why certain aspects of their work were positive or negative. This demonstrates an improvement in students' evaluative expertise over time (Sadler, 2010) as they relate their judgments to the judgments of others: they reflect whether their judgments were appropriate or not, looking for ways to improve future feedback content and wondering what they missed when making their judgments that others did see (Boud, Lawson, & Thompson, 2013). The fact that PF messages in both studies contained informational feedback is important because it means that, at the end of our interventions, the students were able to include the three utmost difficult feedback components in their immediate feedback messages: (1) pointing out when the assessee's performance is not at the requested criteria level, (2) why this is the case, and (3) how the performance can be improved (Hattie & Timperley, 2007). Our findings confirm the importance of practice in PA settings since students' expertise in making valuable judgments on a peers' work improved over time (e.g., Boud & Soler, 2015; Panadero, 2016; Sluijsmans, 2002). However, the explanation of PF quality improvement as the effect of practice might be entangled with a memory effect and a norm development effect over time. More specifically, the use of criteria might have had a substantive positive effect on the quality improvement of the PF over time, as the students may have increasingly remembered the feedback criteria. The repeated use of the rubric criteria might have led to some sort of "norm development" in the classroom. It is, therefore, important for future research to further disentangle these effects.

The results of students' self-reported PF skill improvement in Chapter 5 are in line with their actual PF content quality improvement over time. The results of Chapter 6 on students' self-reported PF skills showed a significant increase during PA sessions 1 and 2; however, stagnation was noted after session 2. The results point out to a ceiling effect of this measure. Future studies could explore students' self-reported PF skill evolution within synchronous settings via a more detailed qualitative approach to explain the noted stagnation. Overall, these findings suggest that all students acknowledged that their involvement in multiple PA sessions led to an improvement in their PF skills. This is an important finding as being conscious of one's active enhancement is considered a key element in the development of PF skills (Boud & Soler, 2015). Additionally, the students' perceived PF skill was measured through a quantitative slider scale as opposed to a detailed qualitative analysis of the actual FB. Hence, future PA studies could primarily focus on performance improvement rather than on the development of evaluative expertise (as was the

case in this dissertation) and also include a more general PF quality score (e.g., Peer Feedback Quality Index by Prins, Sluijsmans, & Kirschner, 2006). Both measures could then be compared by determining the correlation between the actual gain and students' perceived gain over time.

Directions for future research

In this dimension, we focused on important precursors of the learning process, which are likely to be supported through participation in PA; i.e., becoming a skilled self-assessor and thereby improving one's evaluative expertise and Self-Regulated Learning (SRL)skills. As the first step, we investigated the development of students' PF skills and found an overall enhancement in PF content quality in both studied settings. Taking this a step further would necessitate the following: (a) measuring the effect of PA participation on students' sustainable assessment skills and SRL competencies, thus looking beyond the immediate gains from receiving feedback and assessing someone else's work, and (b) investigating the quality of the processes that took place during the oral discussion phase and the mindful cognitive processing of the Socrative feedback reports in our studies.

Measuring the effect of PA practice on students' self- and co-regulating skills

In line with Panadero, Jonsson and Strijbos (2016), we suggest that future research should study the impact of multiple opportunities to engage in assessments as peer assessors on students' self- and co-regulating skills. Currently, there is a lack of evidence showing that student involvement in assessments develops self-regulatory abilities (Brown & Harris, 2016; Gielen et al., 2011). More specifically, the relationship between AfL practices and SRL is more often assumed than explicitly supported by research findings (Panadero et al., 2016). When focusing on the consequential validity of this practice, we would have to take a closer look at how the current PA activity is likely to improve the capacity of students to make effective judgments about their work in subsequent tasks and how the activity would help learners meet potential challenges in future practice settings (Boud & Soler, 2015). Then again, it remains challenging to distinguish between the effect of PA practice and the effect of students' learning experience with this task, which is also likely to have an influence on their capacity to make informed judgments about it (Gielen & De Wever, 2015).

Focus on two-way dialogic FB and mindful cognitive processing

Our exploration of the PF content quality of PF messages might be seen as a first step in disentangling the affordances of PA practice on students' evaluative expertise. The results showed that the students preferred non-anonymous oral feedback in the PA procedure, meaning the

opportunity to exchange their interpretations and negotiate meaning in safe learning contexts was important to them. As such, a two-way dialogic feedback setting can be established (Carless, Salter, Yang, & Lam, 2011). A profound and detailed qualitative analysis (e.g., analyzing a video recording of the oral discussion phase) can be valuable in providing an insight into the development of dialogic feedback processes and the role of the teacher in this phase (see structural dimension). Furthermore, we focused on the supply side of PA, not the demand side, except for the exploration of assessees' perceived PF usefulness, which proved to be highly positive. Analyzing the assessees' mindful processing of the content of the Socrative FB reports is a valuable topic for future research in order to explicitly close the feedback loop. Mindful cognitive processing was recently explored by Bolzer, Strijbos and Fischer (2015) and refers to "how deeply the PF has been cognitively processed and understood" (p.425). Their study showed that eye-tracking methodologies provide valid measures to deduce mindful cognitive processing (e.g., during the reading phase when processing PF, for example, when being confronted with contrasting feedback in the Socrative report).

Structural dimension: Organization and management of PA practices

Overview and discussion of results

In the third dimension, possible structural constraints and solutions for engaging students in dialogic PA processes were examined. In this dissertation, we focused on three components that are frequently concentrated upon in the (re)design of these practices: (a) PA task design, (b) enabling technology, and (c) structure and support.

PA task design

With regard to PA task design, three important conditions are at stake when developing sustainable dialogic PA practices: constructive alignment (Heitink et al., 2016), resemblance to real-life professional tasks, and the nature of the PA activity, which should have an impact on the prospects for the development of evaluative expertise (e.g., Boud & Soler, 2015; Carless, 2015).

First, constructive alignment was assured in our intervention studies through the realization of multiple PA sessions that were a fundamental part of the course design (i.e., 30 to 50% of all classroom meetings), as to allow students to go through a cyclical process of performing a task, assessing task performance, identifying points of improvement, and planning future tasks (Van Merriënboer & Sluijsmans, 2009). Students had the opportunity to prepare and assess each other during group presentations/workshops on course-related topics which constituted the core course content. To ensure engagement over time, a written report had to be handed in at the end of the semester. This two-stage assignment approach is claimed to have potential for supporting sustainable feedback practices as students have the opportunity to include useful peer feedback in their final product so as to close the feedback loop (Carless et al., 2011). In order to prepare students for the PA sessions and recognize their role as novices within a discipline, training and/or scaffolds were used. More specifically, in the study in the secondary education setting (Chapter 5), the students were given an introductory lesson on research design and were actively involved in deciding on the content of the rubric criteria. Additionally, the assessors received guidelines to support them while giving feedback. In Chapter 4 and 6, this training was taken a step further where students got an introductory lesson on the theoretical background of PA approaches in higher education. Furthermore, each group had to design their own rubric on one criterion, and these were then consolidated through group discussion; in a last step, the students were given the opportunity to practice the final rubric with their peers based on examples they

brought themselves. As such, the responsibilities of learning and engaging in assessment practices were shared (Heitink et al., 2016).

Second, the nature of imposed tasks resembled real-life professional tasks within the discipline (Dochy, Segers, & Sluijsmans, 1999). In Chapter 5, the task focused on obtaining research skills and the presentation of research results. Acquiring these skills is important as they prepare secondary education students for future academic tasks at university level (van Ginkel, Gulikers, Biemans, & Mulder, 2016). In Chapter 4 and 6, the task was to create a workshop with a focus on the application of didactical principles. Again, these skills are important for future educational sciences students as giving workshops on instructional methods can be part of their future job content. The focus on these generic and transferable skills in itself contributes to ecological validity of the assessment task. More specifically, it was concerned with practices which met immediate assessment needs (formative needs) whilst not compromising the knowledge, skills, and dispositions required to support lifelong learning activities (sustainability) (Boud, 2000).

Third, with regard to the impact on prospects for the development of evaluative expertise, the analyses of the PF content quality showed an increase in students' PF skills in terms of being able to include difficult feedback components in their immediate feedback messages: (1) pointing out when the assessee's performance is not at the requested criteria level, (2) why this is the case, and (3) how the performance can be improved (Hattie & Timperley, 2007). The PA procedure deliberately contained both a written and oral feedback part. The written feedback reports offered opportunities for unhurried reflection (Yang & Carless, 2013). Furthermore, the oral feedback part offered students the opportunity to engage in dialogue and use arguments to resolve disagreements and internalize feedback. As mentioned in the suggestions for future research in the social-affective dimension, the exploration of the content quality of PF messages must be seen as a first step in disentangling the affordances of PA practice on students' evaluative expertise. A profound and detailed analysis of the verbal phase together with its impact on future tasks is needed.

Enabling technology

In the search for workload-efficient means for establishing synchronous PA practices, the MRT-tool Socrative was highly appreciated for its user-friendly applicability and its immediate written feedback feature and other features such as the projection of summaries of students' rubric scoring input via histograms, well-structured display of peer feedback, and ability to generate feedback reports; these features clearly decrease the in-class time as well as the effort on the part of the teachers to get acquainted with the tool. Of course, the availability of a stable and high-speed

Internet connection is a necessary condition for this activity to run efficiently and effectively. Although, in many cases, it has been noted that technology-enabled assessment might evoke uneasiness and/or unusualness (Struyven & Devesa, 2016), the participants in all studies easily adapted to the proposed assessment setting. The applied technology has proven to be *fit for purpose*: it helps students give feedback to peers in a structured manner on content-related and presentation-related criteria (preparatory guidance), provides a structured display of the results to support the oral discussion phase (in-task guidance), and through its provision of structured feedback reports, it supports the potential uptake of feedback for future performance (performance feedback) (Moscrop & Beaumont, 2017). In sum, this dissertation followed Salomon's (2000, last paragraph) advice to ''let technology show us what can be done, and let educational considerations determine what will be done." We strongly advise future research on technology-enhanced assessment to continue using this facilitative approach.

Structure and support throughout the PA process

In this third component of the structural dimension, we focused on three elements of structure and/or support within synchronous PA settings: (a) a filter-out scaffold for the assessees, (b) anonymity used as a scaffold to create a safe learning environment, and (c) students' preference towards teachers' assessment and feedback.

A filter-out scaffold for the assessees

As discussed in the cognitive dimension, our intervention studies showed that when students are offered PA practice opportunities in combination with rubrics and guiding questions for the assessors, they are more likely to develop expertise for making sound evaluative judgments on a peer's work. Based on the suggestion made in recent work by Tsivitanidou and Constantinou (2016), it was also investigated whether students' evaluative expertise could be stimulated when actively supporting assesees to filter out relevant feedback information they received (Chapter 5). This hypothesis could, however, not be confirmed. Possibly, the effect of filtering out might not have been strong enough on its own and its effect might have been diluted by the other PA scaffolds. For example, the fact that students in both conditions received a Socrative feedback report could have been a sufficient support for an adequate feedback filtering, resulting in an increase in PF quality in the consecutive PA sessions regardless of the filter-out effect. In that sense, the filter-out activity for the assessees could possibly be seen as an example of an over-scripting activity, at least within this specific context and task (Dillenbourg, 2002). Another possible interpretation of the non-significant effect might have been the fact that the filter-out scaffold was given while groups were processing the received PF and that students' PF quality was measured during those situations when they were providing feedback. Thus, the scaffold and

the PF quality in subsequent feedback provisions might have been too loosely connected in time and as a consequence, this may have accounted for the missing effect. In line with the suggestions related to the cognitive dimension, analyzing the assessees' mindful processing of the content of the Socrative FB reports provides a valuable topic for future research as it plausible that this process also triggers the development of students' evaluative expertise.

Anonymity used as a scaffold to create a safe learning environment

In the attempt to strike a balance between the creation of a safe learning environment provided through anonymity and the creation of a rich PF setting (Panadero, 2016), in Chapter 6, the use of anonymity was investigated as a temporary scaffold in order to acknowledge students' possible feelings of insecurity and psychological discomfort and to use it as a facilitator in the generation of critical feedback without fearing social repercussions (e.g., Howard et al., 2010; Vanderhoven, Raes, Montrieux, Rotsaert, & Schellens, 2015). It was found that students initially preferred an anonymous setting, but after experiencing both anonymous and non-anonymous PA sessions, they took a rather neutral stance on the importance of anonymity. The qualitative data, however, showed a more diverse image: half of the students preferred a continued anonymous PA setting both after the first and second non-anonymous setting. The second half of the studied population found anonymity less important after the first or second non-anonymous session. More importantly, the participants could clearly differentiate between their attributed importance towards anonymity and whether a transition approach from an anonymous to a non-anonymous setting was a good approach to evolve towards direct interactive non-anonymous feedback settings, which resemble authentic (work) contexts. This differentiation that students make again favors our suggestion to focus on the creation of a supportive class climate rather than students' individual concerns and stances towards anonymity. It can be expected that the sooner students experience a supportive classroom environment within low-consequence settings (Hattie & Brown, 2008) the sooner the interactions can be focused on providing non-anonymous feedback and guidance.

Students' preferred type of teacher assessment and feedback

Harris and Brown (2013) suggest that the successful implementation of PA practices depends on the teacher's ability to adequately prepare students for this activity. The challenge for the teacher, in this regard, lies in finding the balance between leaving the expert role as a teacher-assessor and passing the process completely into the hands of the students (Bonner, 2016). Previous research indicates that students who are used to a teacher-led assessment can become frustrated (Gielen, Dochy, Onghena, Janssens, & Decuyper, 2007) or experience discomfort (Vanderhoven et al.,

2015) when the teacher's assessment and feedback opportunities are completely replaced by peer assessment and feedback. Because of the specific focus on the role of anonymity within PA (cf. social-affective dimension), students' perceptions were investigated with regard to teacher's assessment and feedback were given anonymously via Socrative (and thus intertwined with the assessment and feedback of peers) and non-anonymously (Chapter 4). The results showed that in both the anonymous and non-anonymous conditions, the participants wished to receive clearly identifiable and explicit teacher assessment and feedback. Teacher's expertise and high perceived objectivity were mentioned as the main reasons behind this preference. Similar to recent findings by Hovardas et al. (2014) and Tsivitanidou et al. (2016), the qualitative data indicate that students tend to use a cross-checking technique between the peer and expert feedback in order to deal with contradictory peer assessment and feedback and filter which feedback they take with them for future performances. It seems that both the teacher's moderation of the oral discussion phase as well as the teacher's feedback serves multiple goals (Xu & Carless, 2016). This will be discussed in the next section.

Directions for future research

After discussing the explored structural challenges in engaging students in dialogic PA processes, two foci for future research will be outlined: (a) teacher-enabling processes and (b) the use of a self-choice anonymity mode.

Teacher-enabling processes

As mentioned in the cognitive dimension, a profound and detailed qualitative analysis can be valuable for providing an insight into the development of dialogic feedback processes and the role of the teacher in this phase. During our interventions, the involved teachers moderated the oral discussion phase by asking reflective questions – this included both content-related input to enforce a shared understanding of the criteria as well as social-affective input, involving acts that build up trust and scale up mutual support between assessors and assessees. However, and this can be considered a limitation of our intervention studies, the effect of these moderating actions was neither measured nor analyzed although the contribution of these actions might be substantive. For example, in a study by van Ginkel, Gulikers, Biemans and Mulder (2015), the added value of a teacher in questioning, intervening, and guiding students in the feedback processes led to higher quality feedback. These authors concluded that implementing PA requires both training of students and ongoing monitoring and feedback regarding the efficacy of students' evaluative efforts. In previous studies exploring the role of the teacher during PA activity, Xu and Carless (2016) outline specific teacher-enabling processes which should be measured in future qualitative studies: (a) cognitive scaffolding which involves strategies to promote students'

disciplinary understanding (e.g., rephrasing and modeling), improve self-regulated capacities (e.g., helping students use the criteria), and provide feedback regarding content quality (e.g., providing suggestions for improvement) and (b) social-affective support involving practices that build up students' trust in teachers and peers (e.g., showing interpersonal caring – Murdock, Stephens, & Grotewiel, 2016), whilst cultivating students' rational attitudes towards critical feedback. These essential processes aim to enhance teacher and student feedback literacy which is a worthwhile path for future research along with the consideration of participants' conceptions of assessment (see social-affective dimension).

Towards a dynamic self-choice anonymity mode

In Chapter 6, the students clearly differentiated between the attributed importance of anonymity and whether a transition approach from anonymous to non-anonymous is seen as a good approach towards direct interactive non-anonymous feedback settings. Furthermore, a decrease in attributed importance of anonymity was noted. As a consequence, a future exploration could introduce a dynamic self-choice anonymity mode, in which students themselves can decide when to give their MRT input (non-) anonymously. In itself, the MRT software does not have this option, but one could offer students the opportunity to use an unidentifiable nickname or number, instead of their real name. As such, within the whole PA procedure, students will get the opportunity to decide to give up their anonymity twice: when giving feedback through the MRT system and when participating in the oral discussion phase. Although there is no previous research on this self-choice mode within synchronous face-to-face PA settings, Yu and Wu (2011) explored a similar approach in an asynchronous online PA setting with primary education students (age 10–11 years). They examined if there were any differential effects among the four different identity revelation modes (three fixed modes: real- name, anonymity, or nickname and one dynamic user self-choice mode) on participants' perceptions of their assessors, their classroom climate, and the learning activity in which they were engaged as evidenced by a self-reported questionnaire. The results showed significantly more favorable perceptual impressions of assessees toward their assessors in a self-choice mode than in the anonymity and nickname modes. This effect could be explained by the presence of strongly negative comments, implying that intense, impulsive, and irrational emotions were present in both the anonymous and nickname groups, which might be a result of the assessor's unrecognizable or less recognizable identity and a decreased sense of individual accountability (Yu & Sung, 2015). Yu and Wu's (2011) result might be specifically related to the age of their student sample. In the intervention studies in this dissertation, hostile or non-normative behavior did not occur. In line with the suggestions made by Cartney (2010), who stated that the teacher also has a quality assurance role in PA, anonymity toward the teacher was not guaranteed in our studies as to enable teacher monitoring of students' FB behavior. This

approach might have prevented the dissemination of hostile PF messages and might have led to the teachers making a significant amount of effort to actively involve the students before, during, and after the PA procedure.

Limitations

The contributions and limitations related to the individual studies were addressed in the previous chapters. In this section, the main overarching strengths and limitations will be addressed.

Limitations regarding the scope of the research

In the intervention studies in this dissertation, the students had to give a PA on the performance (oral presentation or workshop) of a group of peers in a face-to-face classroom context in which immediate rubric scores and PF had to be given via MRT (i.e., intergroup PA task). The specificity of this setting limited not only the potential generalizability of the findings but also the selection of the interpersonal variables which were focused upon. For example, *interdependence*, which refers to the amount of consideration given to a group member as a significant member of the group for completing a group task successfully (Prins, Sluijmans, Kirschner and Strijbos, 2005), or *cooperateviness* were not included, as these interpersonal variables are of significance in intra-group PA activities. The aforementioned variables focus on the group process rather than the task performance, as was the case in this dissertation. Future research studies on the interpersonal nature of PA are encouraged to make a thoughtful selection of the variables to be included in the studies.

In relation to the specificity of the PA setting and the selection of the interpersonal variables, it should be noted that we predominantly focused on instructional design features (e.g., offering anonymity to the assessors, embedment in course design, use of rubrics and guidance questions) in order to lower the interpersonal burden and bias that students can experience when participating in face-to-face PA. An alternative approach could involve identifying statistical solutions in coping with these issues (e.g., new formulae to infer individual scores from group scores – Stegman, Strijbos & Sluijsmans, 2017), although these kinds of solutions would, in our opinion, be predominantly suited for process-oriented PA of collaborative learning (for a detailed discussion, see Strijbos, 2016).

As mentioned in the discussion sections, we mainly focused on the supply side of PA (i.e., the development of evaluative expertise of assessors by offering practice opportunities), rather than the receiver side (i.e., focus on performance improvement of the assessee). This approach was mirrored in the nature of our PA tasks (see structural dimension) as well as the nature of the

used assessment criteria and should be taken into account when interpreting these results. We know from previous research that PA is a complex learning task that requires high-level cognitive processing, and previous studies have shown that if students do not master domain-specific knowledge, having to perform a PA of these domain-specific tasks may hinder their learning and performance (van Zundert, Könings, Sluijsmans, & van Merriënboer, 2012). For this reason, students were acknowledged as novices within the discipline and were expected to assess a peer's work in terms of (a) the feasibility of the research proposal (rather than assessing the peers' knowledge on the research subject) (Chapter 5) or (b) in terms of applying didactical principles (rather than assessing the peers' knowledge on the workshop subject) (Chapter 4 & 6) and (c) in terms of presentation. This approach involves installing a new classroom culture in which a symmetrical relationship between the teacher and learners is proposed (Gielen et al., 2011), implying the suggestion made in the social-affective dimension to position PA practices as teaching-learning or pedagogical interactions (Brown & Harris, 2016). Having different opinions is then enriching as long as peers are transparent about how they relate their judgment to the mutually discussed criteria (Gielen et al., 2011). As such, students are seen as key consumers and producers of the formative assessment information (Andrade, 2010), in which the accuracy of PF messages is defined in terms of their appropriateness to the assessment criteria, rather than them being determined by the subject-related expertise of the teacher (or another expert). Of course, this approach narrows the generalizability of our results and its application in other disciplines. For example, in science education, the provision of FB with scientifically correct content is crucial and when analyzing PF content quality, this accuracy component should also be included in the coding process (e.g., Hovardas, Tsivitanidou, & Zacharia, 2014). Furthermore, the value of our approach is supported by findings by Gielen, Peeters, Dochy, Onghena and Struyven (2010) as they found that justification was superior to the accuracy of comments in having a positive impact on performance. Thus, PF goes beyond a corrective function to one which promotes critical discourse (Gan & Hattie, 2014). The specific framing of our PA activities implies that we need to see PF in the context of negotiating meaning and connecting ideas, rather than providing the "right" answers.

Limitations regarding study sample and methodological issues

The first limitation relates to the small sample sizes included in the studies in Chapter 4, 5, and 6. Therefore, it is advised to substantiate the ecological validity of the obtained research findings by increasing the number of school-based intervention studies in secondary, higher, and primary education, including also more diverse learning populations to establish national and international comparative research findings. Furthermore, equal gender distribution was difficult

to acquire, as the majority of students in educational sciences (Chapter 4 & 6) and human sciences (chapter 5) are female. However, these small scale studies made it possible for the researcher to be present at all times to supervise treatment fidelity.

With regard to the sampling procedure of the participants involved in the intervention study, owing to the required time frame (at least one semester) and intensity of the implemented PA practice, we actively sought for teachers who had already implemented PA in their classroom with multiple PA sessions with previous student cohorts. For example, the responsible teachers in secondary education were working with paper-based PA and were looking for an improvement in their practice through the use of technology, which made their current practice suitable for the research objectives. As intensive cooperation with the teachers was needed, they were aware of the general project aims and conditions. Therefore, it might be possible that the intervention's appeal made some teachers adopt unnoticed strategies that were targeted in the intervention group; for example, during the oral discussion phase or when communicating about the rubric criteria, even with the presence of the researcher during all PA sessions. This is a commonly occurring phenomenon in school-based intervention research as keeping teachers blind to treatment distinctions is a seemingly intractable problem in educational research (Boekaerts & Corno, 2005).

In the intervention studies, the participants' individual characteristics such as general achievement level (Shute, 2008), general assessment and feedback conceptions (Brown, 2008; Brown, Harris, & Harnett, 2012), prior PA experience, and personality (Lakhal, Sévigny, & Frenette, 2013) were not considered. These variables can potentially influence students' perceptions of interpersonal processes and the quality of their provided PF messages. We encourage future research to consider taking into account the aforementioned variables. However, in light of the research objectives and obtained results of this dissertation, we do not expect a strong influence of these variables (e.g., personality), as the general classroom climate seems to be of major importance. Every PA setting is different, with different peers and different tasks; consequently, the interpersonal atmosphere will be unknown and the process of creating a psychologically safe and constructive PA environment has to be repeated in every new PA situation. Therefore, interpersonal perceptions may play a detrimental role in every new peer assessment setting. However, the aforementioned variables, like students' general assessment and feedback conceptions, could be of influence to accelerate the realization of a psychologically safe environment (van Gennip, Segers, & Tillema, 2009). As suggested in the discussion on the social-affective dimension, we encourage future research to focus on how to create these safe environments in daily classroom realities.

With regard to the survey studies in Chapter 2 and 3, we must emphasize that these studies are the first large-sample-based explorations of this under-studied problem, and the models

presented in this dissertation should not be considered universal models. Although we included critical reviews by an expert in formative assessment, pretested our questionnaires, and conducted exploratory and confirmatory factor analyses for these measures, we do recommend testing the external and construct validity of our scales in other samples. This is an important concern as the field of educational assessment is characterized by a lack of clarity around many concepts (Parr & Timperley, 2016). Having a shared language is methodologically essential in research investigations involving student and teacher perceptions. Parr and Timperley (2016) questioned to what extent researcher's categorizations or the language reflected in the items actually aligns with what students and teachers have in their mind when responding to assessment questionnaires. The fact that the occurrence of various patterns and factor structures, depending on the studied population in a certain assessment policy context, can lead to different results is not surprising and has been shown in previous assessment studies (e.g., Bonner, 2016; Brown, Harris, O'Quin, & Lane, 2015). In line with this suggestion, and while we were able to explain the differences in relevant constructs to a large extent even with a reduced number of items, it would be interesting to include broader measures of the same constructs in future studies. The construction of our items was mainly inspired by self-report measures in the few small-scale studies on these issues. Another valuable avenue for future research, in this regard, is to cross (inter)disciplinary borders and look at validated instruments in classroom interaction study and social psychology research (e.g., validated instruments on resistance to peer influence by Sumter, Bokhorst, Steinberg and Westenberg [2009]).

Implications for theory and empirical research

This dissertation was guided by the observation that implementing PA within daily classroom practice is a challenging and intensive undertaking. It involves multiple social and human factors that need to be taken into account because PA does not happen in isolation; it produces thoughts, actions, and emotions as a consequence of the interaction of assessees and assessors, which can have an impact on the quality of the PA process (Panadero, 2016), especially in face-to-face classroom contexts (Latané, 1981; Pope, 2005). The studies focusing on the relation between students' perceptions of interpersonal variables, accuracy and anonymity, and their relation with the perceived educational value of PA are first steps toward the broadening of our understanding of this complex issue. Furthermore, in explicitly recognizing the importance of the teachers' stance and actions in developing a classroom culture that supports the enactment of two-way dialogic PF practices, we have opened up a new line of research by studying teachers' awareness of students' perceptions of these interpersonal processes. We hope that researchers build on our instruments

and results in future work on this topic. As suggested, the alignment of teachers' assessment literacy and teachers' assessment conceptions, particularly in relation to human and social conditions of assessment, should be considered a worthwhile continuation of this research line.

By examining students' PF content quality with an adapted version of Gielen and De Wevers' (2015) content analysis scheme, we have contributed to both our understanding of students' PF quality in PA settings in which PF is given immediately as well as of the applicability of the scheme in synchronous PA settings.

Given the lack of knowledge on how to effectively implement PA practices in which two-way dialogic PF processes can be established in face-to-face classroom environments, the applied PA procedure, including a scoring rubric, written feedback, and oral discussion phase, complied with the important demands claimed in literature (i.e., constructive alignment, Heitink et al., 2016), resembled real-life professional tasks, and demonstrated the impact of PA practice on prospects for the development of evaluative expertise (e.g., Boud & Soler, 2015; Carless, 2015). We thus illustrated a well-considered instructional design of PA that can be applied by researchers in further exploring the development of students' evaluative expertise and/or its effect on students' sustainable assessment skills and SRL competencies. The use of response technology was proven to be useful, and we hope researchers further exploit the possibilities of this technology in this regard (e.g., the proposed dynamic self-choice anonymity mode).

With regard to disentangling the role of anonymity in PA, we add to current literature the finding that students clearly differentiate between the amount of importance they attribute to anonymity of the assessors and whether a transition approach from an anonymous to a non-anonymous mode is valuable with regard to establishing two-way dialogic settings. As described in the discussion of the structural dimension, it is hoped that researchers are stimulated by our findings and suggestions to engage in similar intervention studies and, in particular, explore the use of a dynamic self-choice anonymity mode, in which students themselves can decide when to give up their anonymity as the assessor in face-to-face PA settings.

Implications for practice

The relevance of this dissertation to educational practice consists of the insights gained into the interpersonal processes within PA, the adaptability of the followed PA procedure as well as the enabling function of the used technology.

With regard to the insights that could be gained from studying the interpersonal processes within both the survey and intervention studies, our findings confirm Panadero's (2016)

theoretical work on the social nature of PA: PA does not happen in a vacuum and a shallow implementation of it might do more harm than good. Mitigating the interpersonal variables warrants intensive, repeated, and highly interactive PA tasks in which interpersonal processes are actively monitored because, as mentioned in the discussion of the social-affective dimension, expecting students to be automatically immune to negative interpersonal processes, criticism, or any form of negative interactions symbolizes naïveté and wishful thinking (Brown & Harris, 2016). More specifically, educators are encouraged to discuss the inherent social nature before the start of the actual PA task to raise students' awareness on these issues, recognizing that the act of humans who are required to interact with each other by social policy cannot be taken for granted (Brown & Harris, 2016). For example, our survey study in secondary education indicated the importance of students' awareness level on possible negative interpersonal processes, such as students experiencing fear of disapproval when giving critical feedback. Additionally, in our intervention studies in which students were actively involved in deciding on or creating the rubric criteria and received guiding questions to foster high-quality feedback, an overall positive interpersonal PA climate was established. Also, offering anonymity to assessors has proven to contribute to this overall positive climate. Moreover, a study with a transition approach from anonymous to non-anonymous PA revealed that students made a differentiation between attributed importance towards anonymity and the value of non-anonymous dialogic interactions. We, therefore, encourage teachers to take the option of offering temporal anonymity to assessors along in the discussion with students on the inherent social nature of PA. We agree with Harris and Brown (2013) that there are some classes where, owing to negative classroom dynamics, PA may not be viable from day one – building a safe environment is a matter of utmost concern.

With regard to the role of a teacher during the studied PA interventions, it is advised that students also receive explicit non-anonymous teacher assessment and feedback as they tend to use a cross-checking technique in filtering out relevant PF. Furthermore, throughout the studies, it became clear that only guaranteeing anonymity toward the assessors and not toward the teacher allows the teacher to take on a quality-assurance role in order to actively monitor students' evaluative efforts, whilst cultivating students' rational attitudes toward critical feedback.

The PA procedure used in the intervention studies in this dissertation can be a source of inspiration for teachers. With regard to the nature of the PA task, we predominantly focused on students as novices within a discipline, as no specific prior domain knowledge was required to conduct the PA task, except for a basic understanding of didactical principles (Chapter 4 &6) or quality indicators of research (proposals) (Chapter 5), which were specified in introductory lessons. We, therefore, see high potential in applying our PA procedures to other disciplines or

settings in which appropriateness to assessment criteria of the performance being assessed is valued (e.g., art, music, literature, construction, handcrafting etc.). However, if prior domain knowledge is acquired, the procedure can also be applied to other disciplines (e.g., sciences, mathematics etc.). As mentioned by Hovardas et al. (2014), we recommend extending the teacher quality assurance role (see above) to ensuring scientifically correctness of the delivered PF content.

The application of MRT has proven to facilitate reciprocal feedback processes, so teachers are highly encouraged to use this tool to organize PA practice within their classrooms. Students in our studies revealed to be digital natives who easily found their way through the application and its features. Its applicability for PA can also be a stepping stone to other assessment methods (e.g., self-assessment, reflective lessons etc.) or classroom interaction methods (e.g., activating prior knowledge, exit tickets etc.).

Implications for policy

Concerning the implications for policy, the moderate level of teachers' awareness of students' concerns with regard to interpersonal processes within PA (Chapter 3) warrants new investments in both in- and pre-service assessment education initiatives. In line with recent promising work by Xu and Carless (2016), who focused on connecting the field of educational assessment with teacher education, we support the notion that for teachers to become (peer) assessment literate, future initiatives should involve one-day workshop approaches in which AfL methods are explained as the underlying motivation for the practices and for more fundamental approaches. This learning process for teachers should be a consciousness evoking one in which research evidence on effective formative and sustainable practice is discussed and teachers' (mis)conceptions of assessment and the legitimacy of assessment traditions are questioned. More specifically, with regard to the interpersonal nature of PA, teachers (and teacher educators) need to be exposed to circumstances where they can experience PA for themselves, in order to gain an insight into the effects of the interpersonal processes that are inherent to this assessment method. This can be done in communities of practice in which "accounts of practice" (such as the specific practices studies in in this dissertation) are shared through the process of thoughtful reflection on practice (Hounsell & Zou, 2017). In such communities of practice, recommendations made by empirical studies in authentic settings that are directed to the teacher audience can be discussed and reflected upon (e.g., the possible positive effects of the use of anonymity in face-to-face classroom settings, insights into interpersonal processes within PA, and creating a "meta-social" element in classroom interactions). These accounts of practice seem to have a high degree of

"street credibility" for fellow teachers as they exemplify how compromises can be made between contextual factors external to assessment (e.g., class size, teaching schedule, and accountability demands), student needs, and teachers' conceptions.

We have argued that our findings on the interpersonal nature of PA support the recent recommendations in AfL literature to position all classroom interactions, such as PA practices, as teaching-learning or pedagogical interactions. This framing would potentially free up such interactions from the burdens associated with evaluation and accountability and facilitate a sharper focus on diagnosing the needs and strengths of PA and on providing feedback and guidance for improved instruction and learning (Brown & Harris, 2016). Consequently, this plea positions making evaluative judgments about peers and one's own learning and/or performance as a core-competence in the curriculum. More specifically, it becomes part of the teachable and learnable component of self-regulation (Brown & Harris, 2014). We hope that policy-makers and instructional designers consider this repositioning in future curricular frameworks.

In this dissertation, we had positive experiences with technology-enabled assessment through the application of a response technology via portable devices (laptops, tablets, and smartphones). Hence, we want to appeal to ICT-policy makers, facility managers, and principals to make well-considered investment decisions with regard to stable Wi-Fi infrastructure and Bring Your Own Device policies as to create opportunities for teachers to make *fit for purpose* use of the (sometimes free) available software, without having to face the burden of hardware failures.

References

Andrade, H. L. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. *Handbook of Formative Assessment*, 1–18.

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing Procedures for Implementing Peer Assessment in Large Classes Using an Action Research Process. *Assessment & Evaluation in Higher Education*, *27*(5), 427–441. doi:10.1080/0260293022000009302

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research*, *61*(2), 213–238. doi:10.3102/00346543061002213

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. doi.org/10.1080/0969595980050102

Boekaerts, M., & Corno, L. (2005). Self-Regulation in the Classroom: A Perspective on Assessment and Intervention. *Applied Psychology: An International Review*, *54*(2), 199–231.

Bolzer, M., Strijbos, J. W., & Fischer, F. (2015). Inferring mindful cognitive-processing of peer-feedback via eye-tracking: role of feedback-characteristics, fixation-durations and transitions. *Journal of Computer Assisted Learning*, *31*(5), 422–434. doi:10.1111/jcal.12091

Bonner, S. (2016). Teachers' Perceptions about Assessment: Competing Narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 21–39). New York, NY: Routledge.

Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151–167. doi:10.1080/713695728

Boud, D. (2007). Reframing assessment as if learning were important. *Rethinking Assessment in Higher Education: Learning for the Longer Term*, 14–25.

Boud, D., Lawson, R., & Thompson, D. G. (2013). Assessment & Evaluation in Higher Education Does student engagement in self- assessment calibrate their judgement over time ? *Assessment & Evaluation in Higher Education*, 37–41. doi:10.1080/02602938.2013.769198

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, *38*(6), 698–712. doi:10.1080/02602938.2012.691462

Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 1–14. doi:10.1080/02602938.2015.1018133

Boud, D. (2016). *Current influences on changing assessment: Implications for research to make a difference*. Invited keynote address at EARLI SIG 1 Assessment & Evaluation conference. Munich, Germany.

Brookhart, S. (2011). Tailoring feedback. *The Education Digest*. Retrieved from http://search.proquest.com/openview/f5fee203f67d40a5ae0c17d971580789/1?pq-origsite=gscholar&cbl=25066

Brown, G. T. L. (2008). *Students' conceptions of assessment inventory (SCoA Version VI)*. Auckland, NZ

Brown, G. T. L. (2011). Self-regulation of assessment beliefs and attitudes: a review of the Students' Conceptions of Assessment inventory. *Educational Psychology, vol*(issue) page nos. doi.org/10.1080/01443410.2011.599836

Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self- assessment as a core competency. *Frontline Learning Research*, *3*(1), 22–30. doi:10.14786/flr.v2i1.24

Brown, G. T. L., & Harris, L. R. (2016). *Handbook of Human and Social Conditions in Assessment*. Place: Routledge.

Brown, G. T. L., Harris, L. R., & Harnett, J. (2012). Teacher beliefs about feedback within an assessment for learning environment: Endorsement of improved learning over student well-being. *Teaching and Teacher Education*, *28*(7), 968–978. doi:10.1016/j.tate.2012.05.003

Brown, G. T. L., Harris, L. R., O'Quin, C., & Lane, K. E. (2017). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: identifying and understanding non-invariance. *International Journal of Research & Method in Education*, *40*(1), 66–90. doi:10.1080/1743727X.2015.1070823

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning - a Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. doi: 10.3102/00346543065003245

Carless, D. (2017). Students' Experiences of Assessment for Learning (pp. 113–126). Springer Singapore. doi:10.1007/978-981-10-3045-1_8

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, vol*(issue) page nos. doi:10.1080/03075071003642449

Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education*, *35*(5), 551–564. doi:10.1080/02602931003632381

Cheng, K. H., & Tsai, C. C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology*, *28*, 599–618.

Cowie, B., & Harrison, C. (2016). Classroom Processes that Support Effective Assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 335–350). New York, NY: Routledge.

Deneen, C. C., & Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, *3*(1), 1–14. doi:10.1080/2331186X.2016.1225380

Dillenbourg, P. (2002). *Over-scripting CSCL: The risks of blending collaborative learning with instructional design*. Heerlen, Open Universiteit Nederland. Retrieved from https://telearn.archives-ouvertes.fr/hal-00190230/

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, *69*, 145–186. doi:10.2307/1170673

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education*, *24*, 331–350. doi:10.1080/03075079912331379935

Ecclestone, K., & Pryor, J. (2003). "Learning Careers" or "Assessment Careers"? The Impact of Assessment Systems on Learning. *British Educational Research Journal*, *29*(4), 471–488. doi:10.1080/01411920301849

Fastré, G. M. J., van der Klink, M. R., Sluijsmans, D., & van Merriënboer, J. J. G. (2013). Towards an integrated model for developing sustainable assessment skills. *Assessment & Evaluation in Higher Education*, *38*(5), 611–630. doi:10.1080/02602938.2012.674484

Gan, M. J. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, *42*(6), 861–878. doi:10.1007/s11251-014-9319-4

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, *52*, 315–325. doi:10.1016/j.chb.2015.06.019

Gielen, S., Dochy, F., Onghena, P., Janssens, S., & Decuyper, S. (2007). A complementary role for peer feedback and staff feedback in powerful learning environments. In S. Gielen (Ed.), (pp. 157–199). *Leuven: Katholieke universiteit Leuven, Faculteit Psychologie en Pedagogische Wetenschappen Centrum voor Opleidingsdidactiek*..

Gielen, S., Dochy, F., Onghena, P., Struyven, K., & Smeets, S. (2011). Goals of peer assessment and their associated quality concepts. *Studies in Higher Education*, *36*, 719–735. doi:10.1080/03075071003759037

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, *20*(4), 304–315. doi:10.1016/j.learninstruc.2009.08.007

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, *36*, 101–111. doi:10.1016/j.tate.2013.07.008

Hattie, J. (2016). Foreword. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. IX–XVI). New York, NY: Routledge.

Hattie, J., & Brown, G. (2007). Technology for school-based assessment and assessment for learning: Development principles from New Zealand. *Journal of Educational Technology*, *vol*(issue) page nos.

Hattie, J., & Gan, M. (2011). *Instruction based on feedback*. *Handbook of research on learning and instruction*. Place: Routledge. doi:10.4324/9780203839089.ch13

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. doi:10.3102/003465430298487

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, *38*, 21–27. doi:10.1016/j.stueduc.2012.04.001

Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, *17*, 50–62. doi:10.1016/j.edurev.2015.12.002

Higgins, R., Hartley, P., & Skelton, A. (2001). Getting the Message Across: The problem of communicating assessment feedback. *Teaching in Higher Education*, *6*(2), 269–274. doi:10.1080/13562510120045230

Hill, M. F. (2011). "Getting traction": enablers and barriers to implementing Assessment for Learning in secondary schools. *Assessment in Education: Principles, Policy & Practice*, *18*(4), 347–364. doi:10.1080/0969594X.2011.600247

Hounsell, D., & Zou, T. X. P. (2017). Surfacing and Sharing Advances in Assessment: A Communities-of-Practice Approach (pp. 33–48). Springer Singapore. doi:10.1007/978-981-10-3045-1_3

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers and Education*, *71*, 133–152. doi:10.1016/j.compedu.2013.09.019

Howard, C. D., Barrett, A. F., & Frick, T. W. (2010). Anonymity to Promote Peer Feedback: Pre-Service Teachers' Comments in Asynchronous Computer-Mediated Communication. *Journal of Educational Computing Research*, *43*(1), 89–112. doi:10.2190/EC.43.1.f

Lakhal, S., Sévigny, S., & Frenette, É. (2013). Personality and preference for evaluation methods: A study among business administration students. *Studies in Educational Evaluation*, *39*(2), 103–115. doi:10.1016/j.stueduc.2013.02.002

Latané, B. (1981). The psychology of social impact. *American Psychologist*, *36*(4), 343–356. doi:10.1037/0003-066X.36.4.343

McInerney, D. M., Brown, G. T. L., & Liem, A. D. (2009). *Student Perspectives on Assessment: What Students Can Tell Us about Assessment for Learning*. Retrieved from https://books.google.com/books?hl=nl&lr=&id=9iZy0exq7goC&pgis=1

Moscrop, C., & Beaumont, C. (2017). Technology-Enhanced Assessment Feedback (pp. 193–207). Springer Singapore. doi:10.1007/978-981-10-3045-1_13

Murdock, T. B., Stephens, J. M., & Grotewiel, M. M. (2016). Students Dishonesty in Face of Assessment: Who, Why and What We Can Do About It. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment*. New York, NY: Routledge.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 125-143). Mahwah, NJ: Erlbaum.

Orsmond, P., Merry, S., & Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, *41*, 273–290. doi:10.1080/14703290410001733294

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 247–266). New York, NY: Routledge.

Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: positive experience predicts use. European Journal of Psychology of Education, *32*(1), 133-156. Springer Netherlands. doi:doi.org/10.1007/s10212-015-0282-5

Panadero, E., Jonsson, A., & Strijbos, J.-W. (2016). Scaffolding Self-Regulated Learning Through Self-Assessment and Peer Assessment: Guidelines for Classroom Implementation. In *Assessment for Learning: Meeting the Challenge of Implementation* (pp. 311–326). doi:10.1007/978-3-319-39211-0

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 133–148. doi:10.1080/0969594X.2013.877872

Parr, J. M., & Timperley, H. (2016). Teachers and Assessment: Enhancing Assessment Capability. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 95–108). New York, NY: Routledge.

Pope, N. K. L. (2005). The impact of stress in self-and peer assessment. *Assessment & Evaluation in Higher Education*, *30*, 51–63.

Prins, F. J., Sluijsmans, D. M. A., & Kirschner, P. A. (2006). Feedback for general practitioners in training: Quality, styles, and preferences. *Advances in Health Sciences Education*, *11*(3), 289–303. doi:10.1007/s10459-005-3250-z

Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, *34*(1), 1–20. doi:10.1080/03054980701476386

Regehr, G., & Eva, K. (2006). Self-assessment, Self-direction, and the Self-regulating Professional. *Clinical Orthopaedics and Related Research*, *PAP*, 34–8. doi:10.1097/01.blo.0000224027.85732.b2

Rowe, A. D. (2017). Feelings About Feedback: The Role of Emotions in Assessment for Learning (pp. 159–172). Springer Singapore. doi:10.1007/978-981-10-3045-1_11

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, *35*, 535–550.

Salomon, G. (2000). It's not just the tool, but the educational rationale that counts. Invited keynote address at the 2000 ED-MEDIA Meeting. In *ED-MEDIA Meeting, Montreal, Canada*.

Segers, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation*, *37*(1), 49–54. doi:10.1016/j.stueduc.2011.03.008

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*, 153–189.

Sluijsmans, D. M. A. (2002). *Student involvement in assessment: The training of peer assessment skills.* (Unpublished doctoral dissertation). Open University of the Netherlands, Heerlen.

Stegman, K., Strijbos, J.W., & Sluijsmans, D. (2017). *Inferring individual scores from group scores via peer assessment: Part 2 – A simulation study.* Accepted symposium abstract for the EARLI 2017 Conference. Tampere, Finland.

Strijbos, J.W. (2016). Assessment of collaborative learning (Chapter 17). In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment*. New York, NY: Routledge.

Struyven, K., & Devesa, J. (2016). Students' perceptions of novel forms of assessment: (Chapter 9). In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment*. New York, NY: Routledge.

Sumter, S. R., Bokhorst, C. L., Steinberg, L., & Westenberg, P. M. (2009). The developmental pattern of resistance to peer influence in adolescence: will the teenager ever be able to resist? *Journal of Adolescence*, *32*, 1009–1021. doi:10.1016/j.adolescence.2008.08.010

Topping, K. J. (1998). Peer Assessment between Students in Colleges and Universities. *Review of Educational Research*, *68*(3), 249. doi:10.2307/1170598

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Title of Book* (pp. 55–87). Dordrecht, The Netherlands: Kluwer Academic.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*(4), 339–343. doi:10.1016/j.learninstruc.2009.08.003

Tsivitanidou, O. E., & Constantinou, C. P. (2016). A study of students' heuristics and strategy patterns in web-based reciprocal peer assessment for science learning. *The Internet and Higher Education*, *29*, 12–22. doi:10.1016/j.iheduc.2015.11.002

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, *4*, 41–54. doi:10.1016/j.edurev.2008.11.002

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, *20*, 280–290. doi:10.1016/j.learninstruc.2009.08.010

van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2015). The impact of the feedback source on developing oral presentation competence. *Studies in Higher Education*, 1–15. doi:10.1080/03075079.2015.1117064

van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2016). Fostering oral presentation performance: does the quality of feedback differ when provided by the teacher, peers or peers guided by tutor? *Assessment & Evaluation in Higher Education*, 1–14. doi:10.1080/02602938.2016.1212984

van Zundert, M. J., Könings, K. D., Sluijsmans, D. M. A., & van Merriënboer, J. J. G. (2012). Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously. *Educational Studies*, *38*(5), 541–557. doi:10.1080/03055698.2012.654920

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, *81*, 123–132. doi:10.1016/j.compedu.2014.10.001

Vickerman, P. (2009). Student perspectives on formative peer assessment: an attempt to deepen learning? *Assessment & Evaluation in Higher Education*, *34*, 221–230. doi:10.1080/02602930801955986

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. doi:10.1016/j.tate.2016.05.010

Xu, Y., & Carless, D. (2016). "Only true friends could be cruelly honest": Cognitive scaffolding and social-affective support in teacher feedback literacy. *Assessment & Evaluation in Higher Education*, 1–13. doi:10.1080/02602938.2016.1226759

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, *18*(3), 1–13. doi:10.1080/13562517.2012.719154

Yu, F.-Y., & Sung, S. (2015). A mixed methods approach to the assessor's targeting behavior during online peer assessment: Effects of anonymity and underlying reasons. *Interactive Learning Environments*, 1–18. doi:10.1080/10494820.2015.1041405

Yu, F.-Y., & Wu, C.-P. (2011). Different identity revelation modes in an online peer-assessment learning environment: Effects on perceptions toward assessors, classroom climate and learning activities. *Computers & Education*, *57*(3), 2167–2177. doi:10.1016/j.compedu.2011.05.012

Nederlandstalige samenvatting

Summary in Dutch

De sociale aard van peer assessment in secundair en hoger onderwijs.

Een onderzoek naar studentenpercepties omtrent interpersoonlijke processen en peerfeedbackkwaliteit in anonieme face-to-facecontexten met behulp van Mobiele Responstechnologie

Inleiding

In het onderwijsonderzoek wordt algemeen aangenomen dat assessment een belangrijke drijfveer is in het leerproces van studenten en dat doordachte assessmentpraktijken aanzienlijk kunnen bijdragen tot betekenisvol leren (bv. Black & Wiliam, 1998; Carless, 2017). De recente ontwikkelingen in het onderzoeksveld benadrukken het belang van formatieve benaderingen: dit zijn evaluatieprocessen die studenten informatie verschaffen omtrent de status van hun leerproces, hun leerdoelen en hoe die te bereiken (Sluijsmans, 2014). Andere onderzoekers gaan nog een stap verder. Zij onderschrijven duurzame assessmentpraktijken ('*sustainable assessment*') en stellen dat de doelstellingen van *formative* assessment veelbelovend zijn. Niettemin vinden ze dat deze benadering te weinig inzet op het stimuleren van evaluatievaardigheden die voorbijgaan aan de gerelateerde evaluatieopdracht (bv. Boud & Soler, 2015; Fastré, van der Klink, Sluijsmans, & van Merriënboer, 2013). Dergelijke evaluatievaardigheden zijn nochtans essentieel voor *lifelong learners*: in een snel veranderende wereld met snel veranderende (job)inhouden zullen mensen in staat moeten zijn hun eigen leernoden te identificeren, bij te stellen en ernaar te handelen (bv. Boud & Soler, 2015). Assessment wordt in deze visie opgevat als een kernelement in de onwikkeling van het beoordelingsvermogen, wat beschouwd wordt als noodzakelijk om leerprocessen zelfstandig te kunnen reguleren. Feedback maakt inherent deel uit van deze evaluatievorm en wordt dan ook opgevat als het essentieel leeraspect hierin. Bijgevolg wordt feedback gedefinieerd als een dialogisch proces waarin de student ontvangen informatie doorgrondt met als doel eruit te leren. Met 'dialoog' doelt men op vormen van interactie waarin de actoren samen tot betekenisvol leren komen (Yang & Carless, 2013).

Vanwege de actieve betrokkenheid van de student in het toetsproces wordt peer assessment (PA) gezien als een toetsvorm die nauw aansluit bij zowel een formatieve als een duurzame visie op toetsing (bv. Topping, 2010). Een PA-activiteit wordt in dit proefschrift

gedefinieerd als een activiteit waarin studenten elkaars werk, presentatie of groepsproduct beoordelen. Die beoordeling kan worden uitgedrukt in de vorm van (rubric)scores, mondelinge of geschreven feedback (i.e. peer feedback (PF)) of een combinatie van beide. Onderzoek heeft de voorbije twee decennia aangetoond dat het PA-proces voordelen kan bieden voor zowel de beoordeelde, omdat hij/zij feedback ontvangt, als de beoordelaar. De beoordelaar wordt zich immers meer bewust van zijn/haar eigen prestatie als hij/zij het werk van een medestudent kan afwegen tegen zijn/haar interne representatie van het eigen geleverde werk (bv. Panadero, 2016). In die zin stimuleert PA de groei van het eigen beoordelingsvermogen (Reinholz, 2015). Het implementeren van PA in leeromgevingen is evenwel een veeleisende en intense praktijk. Een PA-activiteit is immers per definitie een interpersoonlijk proces waarbij verschillende gedachten, acties en emoties een rol spelen die het leerpotentieel van deze activiteit kunnen beïnvloeden (Panadero, 2016), in het bijzonder in een face-to-facecontext (Latané, 1981; Pope, 2005). Het is daarom essentieel dat onderwijsonderzoek inzet op het verkennen van effectieve interventies waarin studenten het geven en ontvangen feedback kunnen ervaren en bijgevolg de groei van het beoordelingsvermogen van studenten stimuleren in een leeromgeving waarin ze zich veilig en comfortabel voelen. Hoewel het bestaande onderzoek vaak claimt dat het in zulke situaties wenselijk is om de beoordelaars anonimiteit te bieden (Vickerman, 2009), is hieromtrent vooralsnog weinig onderzoek verricht (Howard, Barrett, & Frick, 2010). Daarom bestudeert dit doctoraat de percepties van studenten omtrent het bieden van anonimiteit aan de beoordelaar en het belang dat zowel studenten als leerkrachten hieraan hechten in relatie tot de inherent sociale aard van PA.

Conceptueel kader en onderzoeksdoelstellingen

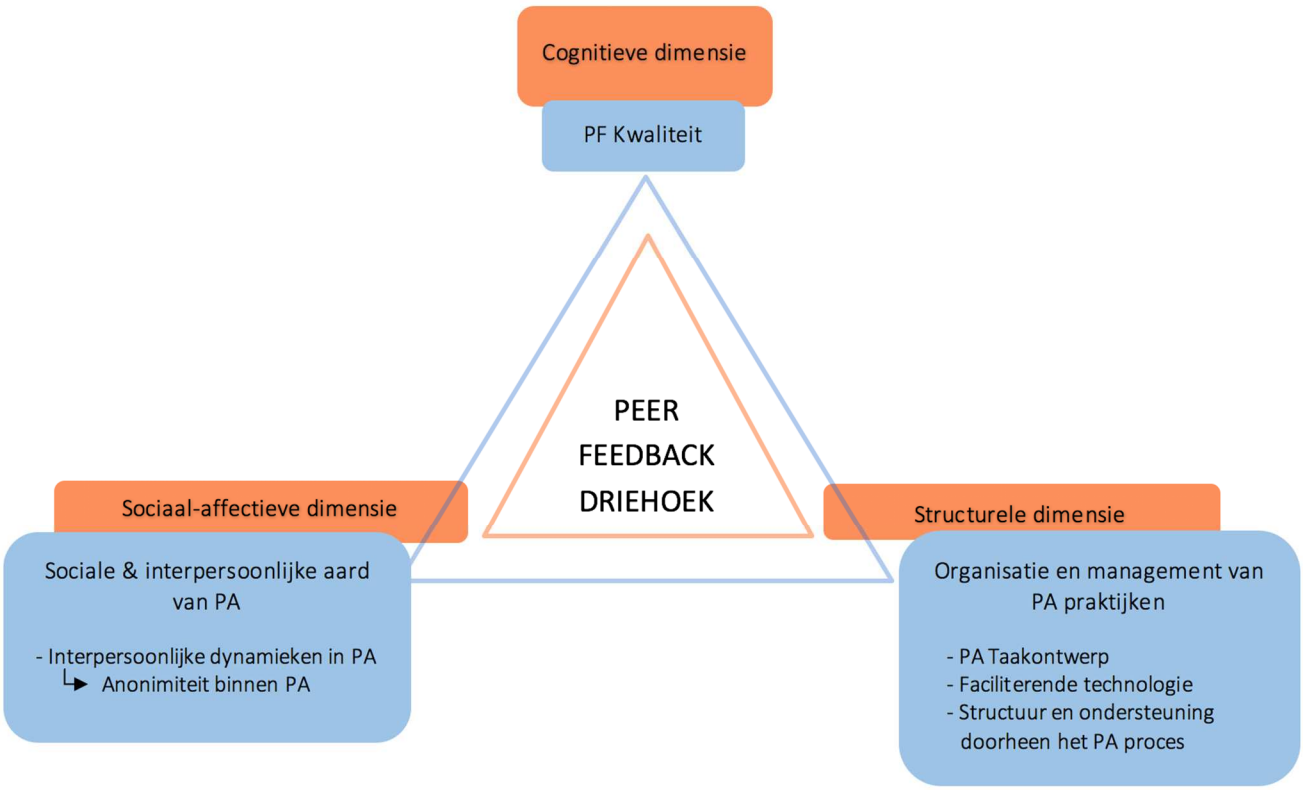
In dit proefschrift hanteren we een aangepaste versie van het conceptuele kader van Yang en Carless (2013) waarin drie dimensies voor de realisering en bestudering van dialogische PA-processen worden onderscheiden (zie figuur 1):

(a) De *sociale en interpersoonlijke aard* van PA – dit behelst de sociaal-affectieve dimensie. We focussen hierbij op sociale en interpersoonlijke condities die de realisatie van waardevolle PA-activiteiten met een PF-component kunnen bevorderen/ verhinderen. Dit werd onderzocht in face-to-face PA-settings waarin studenten onmiddellijk PF geven met behulp van Mobiele Responstechnologie (MRT). Meer specifiek onderzochten we wat de percepties van studenten zijn ten opzichte van (1) interpersoonlijke variabelen (bv. vertrouwen in de evaluatiecapaciteiten van peers, zich psychologisch veilig voelen om deel te nemen aan PA), (2) het belang van anonimiteit en (3) de accuraatheid van het PA-proces in relatie tot de door hen

gepercipieerde educatieve waarde van PA (hoofdstukken 2, 4 & 6). Hoofdstuk 3 onderzoekt in welke mate leerkrachten zich bewust zijn van de bekommernissen die studenten hebben omtrent de bovengenoemde interpersoonlijke processen.

(b) *PF kwaliteit* – dit omvat de cognitieve dimensie. We hebben aan de hand van de aanwezigheid van specifieke structurele componenten (i.e. verificaties en elaboraties) (Gielen & De Wever, 2015) het effect gemeten dat opeenvolgende PA-sessies hebben op de inhoudelijke kwaliteit van de feedbackberichten van studenten (hoofdstuk 5 & 6) .

(c) *Organisatie en management van PA praktijken* – dit behelst de structurele dimensie. We focussen hier op drie componenten: PA-taakontwerp, het inzetten van faciliterende technologie en het bieden van structuur en ondersteuning tijdens het PA-proces (hoofdstukken 4, 5 & 6).



Figuur 3. *Peer Feedback driehoek*

Het inleidende hoofdstuk van dit proefschrift biedt een uitgebreide theoretische achtergrond en een omkadering van de huidige tekorten en uitdagingen in het bestaande onderzoek. Op basis van deze analyse werden de volgende onderzoeksdoelstellingen geformuleerd:

Onderzoeksdoel 1 (OD1):	Het bepalen van het actueel gebruik van PA-praktijken in het secundair onderwijs in Vlaanderen en het onderzoeken van de relatie tussen de PA-concepties van actoren en de inherente sociale aard van deze methode	*Sociaal-affectieve + structurele dimensie*
***Onderzoeksdoel 2 (OD2)**:*	Het onderzoeken van de percepties van studenten omtrent interpersoonlijke variabelen en het gehechte belang aan anonimiteit wanneer betrokken in face-to-face PA gefaciliteerd door MRT.	*Sociaal-affectieve + structurele dimensie*
Onderzoeksdoel 3 (OD3):	Het onderzoeken van de PF-kwaliteit in deze settings.	*Cognitieve dimensie*

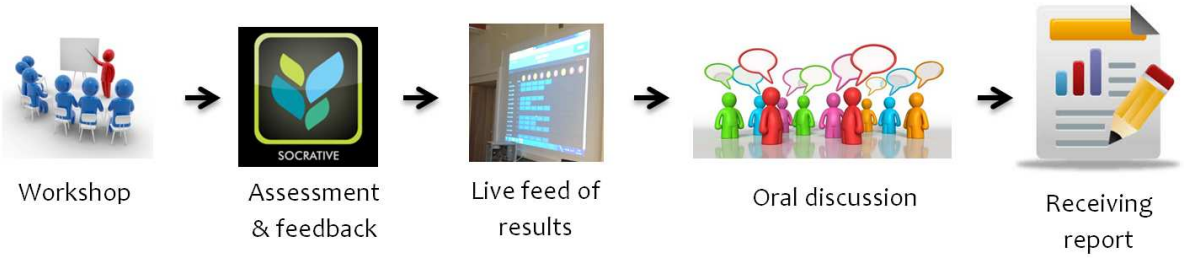
De onderzoeksdoelen werden als volgt bestudeerd in zes hoofdstukken:

In Hoofdstuk 2 wordt in een cross-sectionele surveystudie in het secundair onderwijs het actueel gebruik van PA in kaart gebracht. Daarnaast worden (1) de percepties van studenten (n=3066)omtrent interpersoonlijke variabelen, (2) het belang dat studenten hechten aan anonimiteit en (3) de mate waarin studenten PA accuraat vinden onderzocht, telkens in relatie tot de educatieve waarde die deze studenten aan PA toeschrijven.

In een vergelijkbare studie wordt in hoofdstuk 3 gepeild naar de mate waarin leerkrachten (n=225) in het secundair onderwijs zich bewust zijn van de bekommernissen van studenten omtrent de bovengenoemde interpersoonlijke processen.

Hoofdstuk 4, 5 & 6 behandelen interventiestudies waarin studenten uit het secundair (hoofdstuk 5) en hoger onderwijs (hoofdstuk 4 & 6) deelnemen aan een reciproque (studenten beoordelen elkaar en alle studenten vervullen zowel de rol van beoordelaar als beoordeelde), anonieme (anonimiteit wordt gegarandeerd voor de beoordelaar wanneer hij/zij zijn/haar beoordeling geeft t.a.v. de beoordeelden), synchrone (onmiddellijk oordeel wordt gegeven via mobiele responstechnologie) PA van een groepsopdracht (presentatie of workshop) in een face-to-face setting. De basisprocedure bestaat eruit dat studenten rubricscores en feedback geven via de MRT-applicatie Socrative (www.socrative.com) op een mobiel toestel (laptop, tablet, smartphone) en dat de resultaten anoniem geprojecteerd worden in de klas (zie Figuur 2).

Vervolgens worden de resultaten besproken in een mondelinge discussiefase. Na afloop hiervan worden automatisch gegenereerde en geanonimiseerde feedbackrapporten verstuurd naar het e-mailadres van de beoordeelde. Voor een gedetailleerde uiteenzetting over de gehanteerde ondersteuningsmiddelen (bv. rubrics, gidsvragen) en procedures in elke studie, zie hoofdstukken 4, 5 en 6.



Figuur 2. *Basisprocedure PA-sessie*

Hoofdstuk 4 verkent het gebruik van MRT als een facilitator voor het bewerkstelligen van synchrone anonieme PA en PF in een face-to-face klascontext in het hoger onderwijs. Daarnaast brengt dit hoofdstuk de relatie tussen interpersoonlijke processen en de percepties van studenten omtrent PA in kaart (OD2). Ten derde wordt, in het kader van de specifieke focus op de rol van anonimiteit in dit proefschrift, de voorkeur van studenten omtrent de al dan niet anonieme evaluatie en feedback van de leerkracht in deze setting onderzocht. In deze studie ontvingen 39 studenten hoger onderwijs een groepsopdracht over een toegewezen onderwerp. Zij kregen twee soorten (nl. anonieme en niet-anonieme) evaluatie en feedback van de leerkracht (*within-subjectmanipulatie*). Om een compleet beeld van de interventie te kunnen rapporteren, hebben we gekozen voor een mixed-methodbenadering.

In hoofdstuk 5 wordt het effect onderzocht van de herhaaldelijke beoefening van PA op de PF-kwaliteit bij een groep leerlingen in het secundair onderwijs (n=36) (OD3). Om de evolutie van de kwaliteit van de PF te meten, worden door middel van een inhoudsanalyse in totaal 1561 feedbacksegmenten geanalyseerd, die gespreid zijn over drie meetmomenten. Daarnaast wordt naar aanleiding van eerdere onderzoekresultaten (Tsivitanidou & Constantinou, 2016) een quasi-experimentele studie (experimentele vs. controleconditie) opgezet waarin studenten in één conditie een *scaffold* ontvangen waarin ze actief betrokken worden in het selecteren van relevante informatie uit het ontvangen feedbackrapport. Hierbij wordt de impact van deze *scaffold* op de PF-kwaliteit van studenten bestudeerd.

In de zoektocht naar een balans tussen de creatie van een veilige PA-leeromgeving door middel van het bieden van anonimiteit aan beoordelaars en het creëren van een authentieke dialogische PF-omgeving, hanteren we anonimiteit in hoofdstuk 6 als een tijdelijke *scaffold*, opdat

studenten de tijd krijgen om kritische feedback te leren formuleren, zonder angst voor mogelijke negatieve reacties van de beoordeelden. Daarnaast wordt de evolutie in PF-kwaliteit bij de 46 deelnemers in kaart gebracht. Daarvoor werd een inhoudsanalyse van 4390 feedbacksegmenten uitgevoerd.

Overzicht resultaten

Deze samenvatting biedt een overzicht van de belangrijkste resultaten van dit proefschrift en van de implicaties voor toekomstig onderzoek, de praktijk en het beleid. Deze gegevens worden geplaatst in het kader van de drie dimensies van de *peer feedback driehoek*.

Sociaal-affectieve dimensie: Sociale en interpersoonlijke aard van PA

Zoals aangegeven focust de sociaal-affectieve dimensie op de inherent sociale en interpersoonlijke aard van de PA-activiteiten. De survey studies met leerlingen en leerkrachten uit het secundair onderwijs tonen aan dat zij doorgaans reeds meerdere malen deelgenomen hebben aan PA of meermaals een PA-activiteit hebben georganiseerd. Toch zien we dat er – in tegenstelling tot de aanbevelingen in het bestaande onderzoek – voornamelijk summatief georiënteerde aanpakken worden gerapporteerd waarin scores geven centraal staat. Daarnaast hebben de survey studies de volgende bevindingen opgeleverd:

- In overeenstemming met eerder onderzoek van van Gennip, Segers en Tillema (2010) is de mate van vertrouwen in de eigen evaluatiecapaciteiten en in die van peers een significante voorspeller voor de educatieve waarde die studenten hechten aan PA.
- Met betrekking tot negatieve interpersoonlijke processen (bv. *peer pressure* ondervinden als gevolg van vriendschapsbanden, angst hebben voor de gevolgen bij het geven van kritische feedback), stellen we vast dat wanneer studenten zich goed bewust zijn dat deze processen kunnen voorkomen in PA, dit geen negatief effect heeft op de educatieve waarde die ze toekennen aan PA. Dit zou erop kunnen wijzen dat wanneer studenten zich goed van bewust zijn van de interpersoonlijke processen van de methode, zij zich ook capabel achten om hiermee om te gaan. De resultaten van de leerkrachtenstudie geven aan dat leerkrachten zich slechts in beperkte mate bewust zijn van de bekommernissen die studenten hieromtrent hebben.
- De mate waarin men PA accuraat vindt blijkt zowel bij studenten als bij leerkrachten de belangrijkste voorspeller te zijn voor de aan PA toegekende educatieve waarde.

- Het gebruik van anonieme PA-vormen was over het algemeen laag. Er werd een negatieve relatie gevonden tussen de mate waarin studenten anonimiteit belangrijk vinden en de toegekende educatieve waarde van PA. Met andere woorden, hoe meer studenten belang hechten aan anonimiteit binnen PA, hoe minder positief zij staan t.a.v. PA.

Daarnaast werden de percepties van studenten omtrent interpersoonlijke variabelen in authentieke leercontexten in het hoger onderwijs voor, tijdens en na hun deelname aan meerdere PA-sessies binnen synchrone face-to-face PA-settings onderzocht (hoofdstuk 4 & 6).

Hierbij kwamen we tot volgende bevindingen:

- De resultaten geven aan dat een PA-setting werd gecreëerd waarin studenten zich psychologisch comfortabel voelden om deel te nemen. Bovendien hadden ze een hoge mate van vertrouwen in hun eigen evaluatiecapaciteiten alsook in die van hun peers. De variabele 'vertrouwen' blijkt een positieve voorspeller voor de PA-concepties van studenten.
- Daarnaast werd een hoge unanimiteit vastgesteld m.b.t. de doelen en criteria van de PA-taak. Ook deze variabele bleek een significant positieve voorspeller.
- De meerderheid van de studenten gaf aan dat zij dankzij de geboden anonimiteit in staat waren eerlijke scores en feedback te geven en dus minder beïnvloed werden door vriendschapsbanden. De mate van angst voor mogelijke negatieve repercussies bij het geven van een lage score of kritische feedback was laag en ook hier werd de geboden anonimiteit als verklaring gegeven. De variabele 'angst voor negatieve gevolgen' bleek een significant negatieve voorspeller voor de PA-concepties van studenten. Beide bevindingen bevestigen de theoretische claim dat anonimiteit kan helpen om de interpersoonlijke 'drempels' te verlagen (bv. Yu & Sung, 2015).
- De beschrijvende resultaten in hoofdstuk 6 bevestigen de bevindingen van hoofdstuk 4: de percepties van studenten omtrent positieve interpersoonlijke variabelen (psychologische veiligheid, vertrouwen in de eigen evaluatieve capaciteiten en in die van hun peers, unanimiteit m.b.t. tot doelen en criteria van de PA-taak) waren positief in de anonieme fase. Deze percepties bleven positief wanneer de anonimiteit achterwege werd gelaten. Met betrekking tot negatieve interpersoonlijke variabelen (angst voor negatieve gevolgen van het geven van een lage score of kritische feedback, de invloed van vriendschaps-banden) werd vastgesteld dat deze variabelen in de anonieme fase een lage score kregen; ook in de niet-anonieme sessies werd het belang hiervan laag ingeschat.

Onze bevindingen aangaande de sociaal-affectieve dimensie leiden tot de volgende implicaties en suggesties voor verder onderzoek. Met betrekking tot interpersoonlijke processen kunnen we stellen dat zowel de survey studies als de interventiestudies hebben aangetoond dat interpersoonlijke processen een significante invloed hebben op de educatieve waarde die zowel studenten als leerkrachten hechten aan PA. Onze aanpak van de PA-procedure, waarin we (1) studenten actief betrekken in het opstellen van de criteria voor de rubric, (2) meerdere PA-sessies inbouwen en (3) gidsvragen aanbieden voor het geven van kwaliteitsvolle feedback, heeft tot een algemeen positief PA-klimaat geleid waarin studenten bereid waren actief deel te nemen. Het aanbieden van anonimiteit aan de beoordelaars droeg – al dan niet tijdelijk – bij tot dit positief klimaat. Zoals eerder onderzoek gesuggereerd heeft, blijkt deze ondersteunende omgeving studenten aan te zetten om hun meningen te delen (Harris, Brown, & Harnett, 2014). Dit positieve klimaat blijkt een belangrijke sleutel tot het succes van onze PA-activiteiten te zijn. We zijn er dan ook van overtuigd dat het belangrijk is dat toekomstig onderzoek inzet op het 'metasociale' element in klasinteracties, om zodoende het bewustzijn van studenten omtrent de interpersoonlijke processen binnen PA te laten toenemen (Cowie & Harrison, 2016). We onderschrijven dan ook de suggestie van Brown en Harris (2016) om PA-praktijken te labelen als leraar-leerlinginteracties of pedagogische interacties in plaats van als assessmentpraktijk. Op deze manier zou PA zich kunnen bevrijden van de interpersoonlijke drempels die studenten ermee associëren. Derhalve kan de focus naar de kernwaarde van de activiteit verplaatst worden: feedback geven en richting geven aan toekomstige prestaties.

Instructieverantwoordelijken doen er goed aan om de inherent sociale en interpersoonlijke natuur van PA grondig te bespreken en te kaderen nog voor de aanvang van de eigenlijke PA-activiteit. Hierbij is het belangrijk te erkennen dat de vereiste om feedback te geven aan elkaar volgens algemeen geldende sociale regels geen evidentie is en een leerproces vereist (Brown & Harris, 2016). Daarnaast is het belangrijk om tijdens de PA-activiteiten het sociaal klimaat voortdurend op te volgen. Een klimaat van wederzijds respect en openheid is essentieel, opdat kritische feedback als constructief zou worden beschouwd (Harris & Brown, 2013; Havnes, Smith, Dysthe, & Ludvigsen, 2012; Heitink, Van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016). Het beperkte inzicht hieromtrent dat werd vastgesteld bij leerkrachten in het secundair onderwijs geeft aan dat professionele ontwikkelingsinitiatieven hierop sterker moeten inspelen (bv. 'Hoe bouw je wederzijds vertrouwen tussen studenten op?').

Cognitieve dimensie: PF Kwaliteit

In de cognitieve dimensie van de PF driehoek onderzochten we het effect van de ervaring van opeenvolgende PA-sessies op de inhoudelijke kwaliteit van de feedbackberichten van studenten. Aan de hand van een aangepaste versie van het inhoudsanalyseschema van Gielen en De Wever (2015) werden de feedbackberichten van studenten getoetst op de aanwezigheid van specifieke structurele componenten (i.e. verificaties en elaboraties) om de evolutie van hun beoordelingsvermogen in kaart te brengen (Hattie & Gan, 2011; Narciss, 2008; Shute, 2008). Verificaties handelen over feedbacksegmenten waarin studenten een evaluatief oordeel uitbrengen. Ze geven dus aan of een aspect van een taakuitvoering correct of foutief was. Elaboraties bevatten relevante informatie die verder bouwt op de verificaties om de PF-ontvanger te helpen een soortgelijke toekomstige opdracht te verbeteren of te optimaliseren. Hierbij wordt een onderscheid gemaakt tussen berichten die extra informatie bieden omtrent de geboden verifiërende beoordeling (i.e. informatieve elaboratie) en berichten die de PF-ontvanger een suggestie bieden om met de verifiërende beoordeling actief aan de slag te gaan (i.e. suggestieve elaboratie). De belangrijkste bevindingen omtrent de inhoudsanalyses worden hieronder opgelijst:

- Met betrekking tot de verificaties weten we uit eerder onderzoek dat studenten geneigd zijn om eerder positieve verificaties te geven (Gielen & De Wever, 2015). In onze studies in hoofdstuk 5 en 6 zien we dat over de sessies heen het aantal positieve verificaties stabiel blijft (hoofdstuk 5) of toeneemt (hoofdstuk 6). In die zin worden de bevindingen van Gielen en De Wever (2015) bevestigd. Dit is op zich niet problematisch, zolang ook de andere componenten aanwezig zijn in de feedbackberichten van studenten.
- Wat betreft de negatieve verificaties, zien we in het algemeen – van de begin- tot de eindmeting – een toename van negatieve verificaties. Bovendien stellen we vast dat na de overgang van een anonieme naar een niet-anonieme setting er een kleine daling van de negatieve verificaties is, maar dat deze in de tweede niet-anonieme sessies opnieuw toenemen. Deze evolutie suggereert dat studenten voordeel halen uit het achtereenvolgens ervaren van een anonieme en een niet-anonieme setting, waarin ze de tijd krijgen om geleidelijk aan hun peers niet-anoniem op de negatieve aspecten in hun werk te kunnen wijzen.
- Omtrent de evolutie in het aantal elaboraties constateren we zowel in hoofdstuk 5 als in hoofdstuk 6 een algemene, significante toename van informatieve en suggestieve elaboraties. Meer specifiek is er een daling van informatieve elaboraties tussen de tweede anonieme en de tweede niet-anonieme sessie. Een mogelijke verklaring hiervoor kan zijn, althans voor de

inhoudelijke criteria, dat studenten, door het herhaaldelijk ervaren van de workshops, eerder verkiezen suggestieve elaboraties te geven en zich hiertoe ook in staat voelen, in plaats van informatieve elaboraties.

Deze studies bevestigen de bevinding van eerder onderzoek dat herhaaldelijke PA-sessies positief bijdragen aan de inhoudsgerelateerde PF–kwaliteit en, in het verlengde hiervan, aan het algemeen beoordelingsvermogen van studenten (zie o.a. Boud & Soler, 2015; Panadero, 2016; Sluijsmans, 2002). De studie in het secundair onderwijs (hoofdstuk 5) was evenwel de eerste studie die ook een significante toename van informatieve elaboraties kon aantonen. Het feit dat de PF-berichten in de beide studies informatieve en suggestieve componenten bevatten, stemt overeen met de aanbevelingen uit eerder feedbackonderzoek (Hattie & Gan, 2011). Studenten bleken immers in staat om in een PA-omgeving waarin ze onmiddellijk PF moesten geven ook de moeilijkste componenten toe te voegen: (1) ze geven aan op welke punten de presentatie of workshop niet overeenstemt met de vereiste kwaliteitscriteria, (2) ze geven aan waarom dit het geval is en (3) hoe dit beter kan (Hattie & Timperley, 2007). Daarom focusten we in deze inhoudsanalyse op belangrijke fundamentele vaardigheden die een vereiste zijn om een bekwame zelf-beoordelaar en zelfregulerende lerende te kunnen worden. Verder onderzoek is vereist om het daadwerkelijke effect van deze PA-praktijk op die zelfregulerende vaardigheden te meten. Daarnaast zijn we ervan overtuigd dat er verder onderzoek naar de ontwikkelingen in de mondelinge discussiefase (via een grondige kwalitatieve analyse) en de verwerking van de feedbackrapporten in deze PA-procedure wenselijk is. Ook in deze fase vinden immers belangrijke verwerkings- en leerprocessen plaats die een rol spelen bij de ontwikkeling van het beoordelingsvermogen.

Structurele dimensie: Organisatie en management van PA praktijken

In de derde dimensie onderzochten we drie maatregelen om de realisatie van dialogische PF processen (zie hoofdstuk 1) mogelijk te maken: (1) PA-taakontwerp, (2) faciliterende technologie, en (3) structuur en ondersteuning tijdens het PA-proces. In de onderstaande paragrafen worden de belangrijkste bevindingen beschreven:

- Gezien de actuele beperkte kennisbasis omtrent de implementatie van PA-praktijken waarin dialogische PF-processen kunnen plaatsvinden in face-to-face klascontexten, hebben we in onze PA procedure geprobeerd ondersteuning en *scaffolds* te implementeren zoals ge-suggereerd is in bestaand – zij het voornamelijk theoretisch – onderzoek naar deze praktijken (bv. Carless, 2015). Onze PA-procedure omvatte het gebruik van een scoring rubric die door de studenten zelf of in onderling overleg met de leerkracht was opgesteld (bv. Panadero &

Romero, 2014), gidsvragen om studenten te stimuleren bij het geven van kwalitatieve feedback (bv. Reinholz, 2015) en een geschreven en mondelinge feedbackfase (bv. Nicol, Thomson, & Breslin, 2014). Verder zorgden we ervoor dat de PA-procedures duidelijk ingebed waren in de structuur van de opleidingsonderdeel en dat *constructive alignment* gegarandeerd was (i.e. de afstemming tussen leerdoelen, leer- en toetsactiviteitten)(bv. Heitink et al., 2016; Van Merriënboer & Sluijsmans, 2009). Bovendien bestonden de PA taken uit opdrachten die de studenten ook in reële situaties in authentieke (werk)contexten moeten kunnen uitvoeren, wat bijdraagt tot de ecologische validiteit.

- De MRT-tool Socrative werd zeer positief geëvalueerd door de participanten in de interventie-studie vanwege de gebruiksvriendelijke interface en de mogelijkheid om onmiddellijk (anonieme) feedback te geven via een laptop, tablet en/of smartphone. Voor de instructie-verantwoordelijke zijn de grootste troeven van deze tool de optie om de resultaten van de rubricscores weer te geven via histogrammen, de overzichtelijke weergave van de feedback-berichten en de mogelijkheid om automatische feedbackrapporten te genereren. Een noodzakelijke voorwaarde hiervoor is echter een stabiele en krachtige internetverbinding. We roepen ICT-beleidsmakers, faciliteitmanagers en schoolverantwoordelijken dan ook op weloverwogen keuzes te maken aangaande investeringen in Wifi-infrastructuur en *Bring Your Own Device*-beleid (BYOD), opdat leerkrachten en docenten optimaal gebruik kunnen maken van bestaande (gratis) applicaties in functie van krachtige leerdoelen.
- In hoofdstuk 5 onderzochten we of de actieve begeleiding van studenten bij het selecteren van relevante feedback uit de feedbackrapporten een impact heeft op PF-kwaliteit in de daaropvolgende PA-sessies. Deze hypothese kon echter niet bevestigd worden. Een mogelijke verklaring hiervoor is dat het effect van deze filter-out *scaffold* beïnvloed werd door de bijdrage van andere gehanteerde hulpmiddelen (cf. over-scripting activiteit - Dillenbourg, 2002). Zo kan het zijn dat het feit dat beide studenten gestructureerde feedbackrapporten ontvingen op zich voldoende ondersteuning bood, wat bijgevolg resulteerde in de PF-kwaliteitstoename in de daaropvolgende PA-sessies.
- In hoofdstuk 6 werd aangetoond dat het feit dat studenten anonimiteit als beoordelaar belangrijk vinden niet verhindert dat zij tegelijkertijd begrijpen dat de afbouw van anonimiteit in een PA-activiteit wenselijk is om toe te werken naar authentieke interactieve PF-dialogen. Bovendien werd in de loop van deze interventieaanpak een daling vastgesteld in het belang dat studenten hechtten aan anonimiteit. Op basis van de bovenstaande bevindingen adviseren we instructieverantwoordelijken de optie te overwegen in een beginfase de anonimiteit van beoordelaars te garanderen met het oog op de creatie van een veilig leerklimaat én dat deze optie wordt meegenomen in de discussie over de inherent interpersoonlijke aard van PA, zoals beschreven in de sociaal-affectieve dimensie.

- Met betrekking tot de rol van de instructieverantwoordelijke in deze PA-procedure werd het duidelijk dat studenten expliciete niet-anonieme feedback van de instructieverantwoordelijke verkiezen. Onze analyses brachten aan het licht dat studenten een *cross-checking* techniek hanteren, waarbij de feedback van de instructieverantwoordelijke als een referentiepunt wordt gehanteerd bij de selectie van de PF die wordt meegenomen met het oog op toekomstige prestaties (Hovardas, Tsivitanidou, & Zacharia, 2014). Daarnaast werd het tijdens de interventiestudies duidelijk dat het wenselijk is dat de feedback van beoordelaar niet anoniem is voor de instructieverantwoordelijke opdat een actieve monitoring mogelijk wordt van de inspanningen die studenten leveren om het werk van hun peers te beoordelen.

Referenties

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. doi:10.1080/0969595980050102

Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 1–14. doi:10.1080/02602938.2015.1018133

Brown, G. T. L., & Harris, L. R. (2016). *Handbook of Human and Social Conditions in Assessment*. New York, NY: Routledge.

Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, *69*(6), 963–976. doi:10.1007/s10734-014-9816-z

Carless D. (2017). *Scaling up Assessment for Learning in Higher Education* (Vol. 5). Springer Verlag, Singapore. doi:10.1007/978-981-10-3045-1

Cowie, B., & Harrison, C. (2016). Classroom Processes that Support Effective Assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 335–350). New York, NY: Routledge.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. Heerlen, Open Universiteit Nederland.

Fastré, G. M. J., van der Klink, M. R., Sluijsmans, D., & van Merriënboer, J. J. G. (2013). Towards an integrated model for developing sustainable assessment skills. *Assessment & Evaluation in Higher Education*, *38*(5), 611–630. doi:10.1080/02602938.2012.674484

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, *52*, 315–325. doi:10.1016/j.chb.2015.06.019

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, *36*, 101–111. doi:10.1016/j.tate.2013.07.008

Harris, L. R., Brown, G. T. L., & Harnett, J. A. (2014). Understanding classroom feedback practices: A study of New Zealand student experiences, perceptions, and emotional responses. *Educational Assessment, Evaluation and Accountability*, *26*(2), 107–133. doi:10.1007/s11092-013-9187-5

Hattie, J., & Gan, M. (2011). *Instruction based on feedback*. *Handbook of research on learning and instruction*. Routledge. doi:10.4324/9780203839089.ch13

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. doi:10.3102/003465430298487

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, *38*, 21–27. doi:10.1016/j.stueduc.2012.04.001

Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, *17*, 50–62. doi:10.1016/j.edurev.2015.12.002

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers and Education*, *71*, 133–152. doi:10.1016/j.compedu.2013.09.019

Howard, C. D., Barrett, A. F., & Frick, T. W. (2010). Anonymity to Promote Peer Feedback: Pre-Service Teachers' Comments in Asynchronous Computer-Mediated Communication. *Journal of Educational Computing Research*, *43*(1), 89–112. doi:10.2190/EC.43.1.f

Latané, B. (1981). The psychology of social impact. *American Psychologist*, *36*(4), 343–356. doi:10.1037/0003-066X.36.4.343

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), Handbook of research on educational communications and technology (pp. 125-143). Mahwah, NJ: Erlbaum

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102–122. doi:10.1080/02602938.2013.795518

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 247–266). New York, NY: Routledge.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 133–148. doi:10.1080/0969594X.2013.877872

Pope*, N. K. L. (2005). The impact of stress in self-and peer assessment. *Assessment & Evaluation in Higher Education*, *30*, 51–63.

Reinholz, D. (2015). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 1–15. doi:10.1080/02602938.2015.1008982

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*, 153–189.

Sluijsmans, D. (2014). *Toetsen met leerwaarde: Een reviewstudie naar de effectieve kenmerken van formatief toetsen*.

Sluijsmans, D. M. A. (2002). Student involvment in assessment: The training of peer assessment skills. (Unpublished doctoral dissertation). Open University of the Netherlands, Heerlen.

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, *20*(4), 339–343. doi:10.1016/j.learninstruc.2009.08.003

Tsivitanidou, O. E., & Constantinou, C. P. (2016). A study of students' heuristics and strategy patterns in web-based reciprocal peer assessment for science learning. *The Internet and Higher Education*, *29*, 12–22. doi:10.1016/j.iheduc.2015.11.002

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, *20*, 280–290. doi:10.1016/j.learninstruc.2009.08.010

Van Merriënboer, J. J. G., & Sluijsmans, D. M. A. (2009). Toward a synthesis of cognitive load theory, four-component instructional design, and self-directed learning. *Educational Psychology Review*, *21*(1), 55–66. doi:10.1007/s10648-008-9092-5

Vickerman, P. (2009). Student perspectives on formative peer assessment: an attempt to deepen learning? *Assessment & Evaluation in Higher Education*, *34*, 221–230. doi:10.1080/02602930801955986

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, *18*(3), 1–13. doi:10.1080/13562517.2012.719154

Yu, F.-Y., & Sung, S. (2015). A mixed methods approach to the assessor's targeting behavior during online peer assessment: effects of anonymity and underlying reasons. *Interactive Learning Environments*, 1–18. doi:10.1080/10494820.2015.1041405

Academic output

Academic output

Journals (A1)

Rotsaert, T., Panadero, E., Estrada, E., & Schellens, T. (2017). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. *Studies in Educational Evaluation*, 53, 29-40. doi: 10.1016/j.stueduc.2017.02.003

Rotsaert, T., Schellens, T., & Panadero, E. (submitted). How do teachers perceive the educational value of peer assessment in relation to its social nature? Manuscript under review for publication in *Studies in Educational Evaluation.*

Rotsaert, T., Schellens, T., Raes, A., & Van Hoey, S. (submitted). Use of Mobile Response Technology for Anonymous Peer Assessment and Feedback: Exploring the Interrelatedness of Interpersonal Variables and Students' Preferred Type of Teacher Assessment. Manuscript under review for publication in *Journal of Educational Computing Research.*

Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2017). "Now you know what you're doing right and wrong"! Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education*. doi: 10.1007/s10212-017-0329-x .

Rotsaert, T., Panadero, E., & Schellens, T. (submitted). Anonymity as an instructional scaffold in peer assessment: its effects on peer feedback quality and evolutions in students' perceptions about peer assessment skills. Manuscript under minor revision for publication in a forthcoming special issue in the *European Journal of Psychology of Education.*

Conference contributions

Rotsaert, T., & Schellens, T. (2014). *Anonymity within face-to-face peer assessment: exploring the role of the teacher*. Paper presented at the 7th EARLI SIG1: Assessment and Evaluation, Madrid, Spain, 27 – 29 August 2014 .

Rotsaert, T., & Schellens, T. (2015). *The interpersonal nature of interactions in peer assessment: focus on anonymity and students' personality.* Paper presented at the 16th Biennial EARLI conference, Limassol, Cyprus, 25 – 29 August 2015.

Rotsaert, T., Raes, A., & Schellens, T. (2015). *Anonymous peer assessment through mobile response technology in higher education: effects of interpersonal variables and students' preferred type of teacher guidance*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Chicago, USA, 16 – 20 April 2015.

Rotsaert, T., & Schellens, T. (2015). *Anonieme face-to-face peer assessment en feedback via Mobile Response Technology: een verkenning van interpersoonlijke variabelen en twee types docentbegeleiding.* Paper presented at the Onderwijs Research Dagen (ORD), Leiden, The Netherlands, 17 – 19 June 2015.

Rotsaert, T., & Schellens, T. (2016). *De evolutie van studentenpercepties ten aanzien van anonimiteit in peer assessment: een interventiestudie in het hoger onderwijs.* Paper presented at the Onderwijs Research Dagen (ORD), Rotterdam, The Netherlands, 22-25 May 2016.

Rotsaert, T., & Schellens, T. (2016). *Exploring students' conceptions towards fading anonymity over time. ICLS conference abstract.* Poster presented at the International Conference of the Learning Sciences (ICLS), Singapore, 21-23 June 2016.

Rotsaert, T., & Schellens, T. (2016). *Exploring students' peer feedback quality in an anonymous synchronous peer assessment setting in Secondary Education*. Paper presented at the 8th Biennial Conference of EARLI SIG 1: Assessment & Evaluation, Munich, Germany, 22-24 August 2016.

Data storage fact sheets

% Data Storage Fact Sheet 1

% Name/identifier study: Chapter 2

% Author: Tijs Rotsaert

% Date: March 20, 2017

1. Contact details

==========

1a. Main researcher

- name: Tijs Rotsaert
- address: Henri Dunantlaan 2, 9000 Ghent, Belgium
- e-mail: Tijs.Rotsaert@UGent.be

1b. Responsible Staff Member (ZAP)

- name: Tammy Schellens (promotor PhD research)
- address: Henri Dunantlaan 2, 9000 Ghent, Belgium
- e-mail: Tammy.Schellens@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies

==========

* Reference of the publication in which the datasets are reported:
Rotsaert, T., Panadero, E., Estrada E. & Schellens T. (2017). How do students perceive the educational value of peer assessment in relations to its social nature? A survey study in Flanders. *Studies in Educational Evaluation. doi*: 10.1016/j.stueduc.2017.02.003.

* Which datasets in that publication does this sheet apply to?:
The sheet applies to all the data used in the publication

3. Information about the files that have been stored

==========

3a. Raw data

---

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:

* On which platform are the raw data stored?

1. Student survey data
   - [X] researcher PC
   - [ ] research group file server
   - [X] other (specify): The raw data is stored in the archive room of the department of Educational studies at PP06.

* Who has direct access to the raw data (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [X] all members of the research group
- [ ] all members of UGent
- [] other (specify):

3b. Other files
---
* Which other files have been stored?
- [X] file(s) describing the transition from raw data to reported results. Specify: SPSS-syntax files were stored.
- [X] file(s) containing processed data. Specify: Survey data was processed and cleaned in SPSS.
- [X] file(s) containing analyses. Specify: MPlus output and models + SPSS output (i.e. results of EFA).
- [] files(s) containing information about informed consent
- [ ] a file specifying legal and ethical provisions
- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
- [ ] other files. Specify: ...

* On which platform are these other files stored?
- [X] individual PC
- [X] research group file server
- [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [ ] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction

==========

* Have the results been reproduced independently?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):
  - name:
  - address:
  - affiliation:
  - e-mail:

% Data Storage Fact Sheet 2
% Name/identifier study: Chapter 3
% Author: Tijs Rotsaert
% Date: March 20, 2017

1. Contact details
==========

1a. Main researcher

- name: Tijs Rotsaert
- address: Henri Dunantlaan 2, 9000 Ghent, Belgium
- e-mail: Tijs.Rotsaert@UGent.be

1b. Responsible Staff Member (ZAP)

- name: Tammy Schellens (promotor PhD research)
- address: Henri Dunantlaan 2, 9000 Ghent, Belgium
- e-mail: Tammy.Schellens@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==========
* Reference of the publication in which the datasets are reported:
Rotsaert, T., Schellens, T. & Panadero, E. (submitted). How do teachers perceive the educational value of peer assessment in relations to its social nature? *Studies in Educational Evaluation.*

* Which datasets in that publication does this sheet apply to?:
The sheet applies to all the data used in the publication

3. Information about the files that have been stored
==========
3a. Raw data
---

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:

* On which platform are the raw data stored?

1. Teacher survey data
   - [X] researcher PC
   - [ ] research group file server
   - [X] other (specify): The raw data is stored in the archive room of the department of Educational studies at PP06.

* Who has direct access to the raw data (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [X] all members of the research group
- [ ] all members of UGent
- [] other (specify):

3b. Other files

---

* Which other files have been stored?
- [X] file(s) describing the transition from raw data to reported results. Specify: SPSS-syntax files were stored.
- [X] file(s) containing processed data. Specify: Survey data was processed and cleaned in SPSS.
- [X] file(s) containing analyses. Specify: MPlus output and models + SPSS output (i.e. results of EFA).
- [] files(s) containing information about informed consent
- [ ] a file specifying legal and ethical provisions
- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
- [ ] other files. Specify: ...

* On which platform are these other files stored?
- [X] individual PC
- [X] research group file server
- [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [ ] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction

==========

* Have the results been reproduced independently?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):
  - name:
  - address:
  - affiliation:
  - e-mail:

% Data Storage Fact Sheet 3
% Name/identifier study: Chapter 4
% Author: Tijs Rotsaert
% Date: March 20, 2017

1. Contact details
==========

1a. Main researcher

- name: Tijs Rotsaert
- address: Henri Dunantlaan 2, 9000 Ghent, Belgium
- e-mail: Tijs.Rotsaert@UGent.be

1b. Responsible Staff Member (ZAP)

- name: Tammy Schellens (promotor PhD research)
- address: Henri Dunantlaan 2, 9000 Ghent, Belgium
- e-mail: Tammy.Schellens@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==========
* Reference of the publication in which the datasets are reported:
Rotsaert, T., Schellens, T., Raes, A. & Van Hoey, S. (submitted). Using Mobile Response Technology for Anonymous Peer Assessment and Feedback: the Interrelatedness of Interpersonal Variables and Students' Preferred Type of Teacher Assessment. *Journal of Educational Computing Research.*

* Which datasets in that publication does this sheet apply to?:
The sheet applies to all the data used in the publication

3. Information about the files that have been stored
==========
3a. Raw data
---

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

1. Teacher survey data
   - [X] researcher PC
   - [ ] research group file server
   - [X] other (specify): The raw data is stored in the archive room of the department of Educational studies at PP06.

* Who has direct access to the raw data (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [X] all members of the research group
- [ ] all members of UGent
- [] other (specify):

3b. Other files

---

* Which other files have been stored?
- [X] file(s) describing the transition from raw data to reported results. Specify: SPSS-syntax files were stored.
- [X] file(s) containing processed data. Specify: The data was cleaned and aggregated in SPSS for Analysis + student reflection notes with open-ended questions + R-input files
- [X] file(s) containing analyses. Specify: SPSS output (i.e. pre-post analysis) + qualitative thematic data analyses files (Excel)+ R output files path analysis
- [] files(s) containing information about informed consent
- [ ] a file specifying legal and ethical provisions
- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
- [ ] other files. Specify: ...

* On which platform are these other files stored?
- [X] individual PC
- [X] research group file server
- [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [ ] all members of the research group

- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction
==========
* Have the results been reproduced independently?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):
  - name:
  - address:
  - affiliation:
  - e-mail:

% Data Storage Fact Sheet 4
% Name/identifier study: Chapter 5
% Author: Tijs Rotsaert
% Date: March 20, 2017

1. Contact details
==========

1a. Main researcher

- name: Tijs Rotsaert
- address: Henri Dunantlaan 2, 9000 Ghent, Belgium
- e-mail: Tijs.Rotsaert@UGent.be

1b. Responsible Staff Member (ZAP)

- name: Tammy Schellens (promotor PhD research)
- address: Henri Dunantlaan 2, 9000 Ghent, Belgium
- e-mail: Tammy.Schellens@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==========
* Reference of the publication in which the datasets are reported:
Rotsaert, T., Panadero, E., Schellens T., & Raes, A. (2017). "Now you know what you're doing right and wrong!" Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education.* doi:10.1007/s10212-017-0329-x

* Which datasets in that publication does this sheet apply to?:
The sheet applies to all the data used in the publication

3. Information about the files that have been stored
==========
3a. Raw data
---

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:

* On which platform are the raw data stored?

1. Teacher survey data
   - [X] researcher PC
   - [ ] research group file server
   - [ ] other (specify):

* Who has direct access to the raw data (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [X] all members of the research group
- [ ] all members of UGent
- [] other (specify):

3b. Other files

---

* Which other files have been stored?
- [X] file(s) describing the transition from raw data to reported results. Specify: SPSS-syntax files were stored + Socrative feedback reports with peer feedback messages and rubric scores
- [X] file(s) containing processed data. Specify: The data was cleaned and aggregated in SPSS for Analysis + coded qualitative Socrative feedback reports with peer feedback messages and rubric scores were quantified and saved in SPSS
- [X] file(s) containing analyses. Specify: SPSS output (i.e. pre-post analysis)
- [] files(s) containing information about informed consent
- [ ] a file specifying legal and ethical provisions
- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
- [ ] other files. Specify: ...

* On which platform are these other files stored?
- [X] individual PC
- [X] research group file server
- [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [ ] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction

==========

* Have the results been reproduced independently?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):
  - name:
  - address:
  - affiliation:
  - e-mail:

% Data Storage Fact Sheet 5
% Name/identifier study: Chapter 6
% Author: Tijs Rotsaert
% Date: March 20, 2017

1. Contact details
==========

1a. Main researcher

- name: Tijs Rotsaert
- address: Henri Dunantlaan 2, 9000 Ghent, Belgium
- e-mail: Tijs.Rotsaert@UGent.be

1b. Responsible Staff Member (ZAP)

- name: Tammy Schellens (promotor PhD research)
- address: Henri Dunantlaan 2, 9000 Ghent, Belgium
- e-mail: Tammy.Schellens@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies
==========
* Reference of the publication in which the datasets are reported:
Rotsaert, T., Panadero, E., & Schellens T. (submitted). Anonymity as an instructional scaffold in peer assessment: its effect on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education.*

* Which datasets in that publication does this sheet apply to?:
The sheet applies to all the data used in the publication

3. Information about the files that have been stored
==========
3a. Raw data
---

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO
If NO, please justify:

* On which platform are the raw data stored?

1. Teacher survey data
   - [X] researcher PC
   - [ ] research group file server
   - [ ] other (specify):

* Who has direct access to the raw data (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [X] all members of the research group
- [ ] all members of UGent
- [] other (specify):

3b. Other files

---

* Which other files have been stored?
- [X] file(s) describing the transition from raw data to reported results. Specify: SPSS-syntax files were stored + Socrative feedback reports with peer feedback messages and rubric scores
- [X] file(s) containing processed data. Specify: The data was cleaned and aggregated in SPSS for Analysis + coded qualitative Socrative feedback reports with peer feedback messages and rubric scores were quantified and saved in SPSS
- [X] file(s) containing analyses. Specify: SPSS output (i.e. pre-post analysis)
- [] files(s) containing information about informed consent
- [ ] a file specifying legal and ethical provisions
- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
- [ ] other files. Specify: ...

* On which platform are these other files stored?
- [X] individual PC
- [X] research group file server
- [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [ ] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction

==========

* Have the results been reproduced independently?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):
  - name:
  - address:
  - affiliation:
  - e-mail: