



Peer assessment use, its social nature challenges and perceived educational value: A teachers' survey study



Tijs Rotsaert^{a,*}, Ernesto Panadero^b, Tammy Schellens^a

^a Department of Educational Studies, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium

^b Departamento de Psicología Evolutiva y de la Educación, Universidad Autónoma de Madrid, Madrid, Spain

ARTICLE INFO

Keywords:

Teachers
Peer assessment
Interpersonal processes
Anonymity
Structural equation modeling

ABSTRACT

The implementation of peer assessment (PA) in the classroom faces considerable interpersonal challenges. In this quantitative survey study ($N = 225$) we focus on the current use and format of PA among secondary education teachers in Flanders and explore teachers' awareness of these interpersonal challenges. We validated an instrument for measuring teachers' awareness which was then used to investigate how this awareness level relates to their conceptions of the educational value of PA. SEM results show that teachers are slightly to 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. This study illustrates that teachers' perceived accuracy of PA is a major predictor of their belief in its educational value and opens up a new avenue for research on teachers' awareness of interpersonal processes in PA.

1. Introduction

Peer assessment (PA) has been shown to have positive effects on students' motivation and engagement in learning (Topping, 2003). In PA, peers use one another as a resource, both by sharing ideas and evaluating the ideas of others, and by providing feedback, 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, such as more enhanced learning (Dochy, Segers, & Sluijsmans, 1999) and training of skills to assess high-quality work in relation to specified criteria (Reinholz, 2015; Topping, 1998), and can be conceived of as a learning tool due to the active involvement of the learner in the learning process (Harris & Brown, 2016; Panadero & Brown, 2017; Topping, 2010). Furthermore, 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 peers (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, resulting in unfair PA or refusal to participate (Raes, Vanderhoven, & Schellens, 2013). Until recently, the social nature (referring to the fact

that PA is per definition an interpersonal process) 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, 2009). 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. The teacher's perspective and actions are critical in developing a culture of classroom assessment that supports sharing ideas beyond individual differences (Cowie & Harrison, 2016; Harris & Brown, 2013). Studying teachers' conceptions of PA is important at a time when the innovation of assessment practices is on the educational agenda (Brown & Harris, 2016), as we know from previous research that conceptions predict their classroom practices (Panadero & Brown, 2017; Rubie-Davies, Flint, & McDonald, 2012).

1.1. Teachers' PA conceptions

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 (e.g. 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

* Corresponding author at: Department of Educational Studies, Ghent University, Henri Dunantlaan 2, BE9000 Ghent, Belgium.

E-mail address: tijs.rotsaert@ugent.be (T. Rotsaert).

PA, they sometimes struggle with its inherent difficulties (e.g. students' possible lack of objectivity) and positive previous experience is one of the main predictors of PA implementation. Interestingly, primary and secondary school teachers reported higher levels of PA implementation and certainty about its educational value than 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). The aforementioned studies build upon Ajzen's (2005) model of planned behavior which suggests that personal beliefs shape one's own behavior, and thus help us to better understand the effect of teachers' beliefs on the self-reported appreciation and use of this assessment practice. However, these studies did not focus on the human and social conditions that can stimulate and/or impede the enactment of valuable PA practices. 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 teachers hold. 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.). 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 PA (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. Currently, however, little is known about teachers' awareness of students' concerns about the interpersonal dynamics in PA, including the lack of instruments on how to measure this. A first step is to develop an adequate instrument and explore teachers' awareness regarding students' perceptions of interpersonal processes involved in PA. The next section describes the existing evidence on students' perceptions of PA interpersonal processes, which provides a starting point to make the transfer to the teachers' point of view.

1.2. Peer assessment: an interpersonal process

PA is fundamentally an interpersonal process as it generates thoughts, actions, motivational outcomes, and emotions for both assesses and assessors (Panadero, 2016). Attention to social and human conditions is thus needed because well-implemented PA should decrease affective threats, ensure accuracy, and lead to positive learning outcomes (Harris & Brown, 2016; Panadero & Brown, 2017; Topping, 2010). In a recent survey study by Rotsaert, Panadero, Estrada, and Schellens (2017), students' perceptions of the educational value of PA in relation to interpersonal variables, anonymity, and accuracy were investigated. As outlined in the introduction, these specific factors will be investigated from a teacher's perspective in the current study, as they can have a predictive value for teachers' classroom practices and will allow us to deepen our understanding of this so-far unexplored area in PA research. The following paragraphs briefly describe our current knowledge on six frequently referred to interpersonal variables and their relevance for the current study (Panadero, 2016): (1) *Friendship marking*; (2) *Fear of disapproval*; (3) *Psychological safety*; (4) *Value congruency*; (5) *Trust in oneself as an assessor*; and (6) *Trust in the other as an assessor*.

(1) *Friendship marking*. 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). From a teachers' point of view it is important to study whether teachers have considered the effect of

friendship bonds on the outcomes of PA activity, so as to be able (in a second stage) to provide adequate social-affective support (Murdock, Stephens, & Grotewiel, 2016). (2) *Fear of disapproval* refers to the assessor's fear of negative comments from the assessee if they give them a low score or negative 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 PA activity exerted a positive effect on students' perceptions of the educational value of PA (Rotsaert et al., 2017a). In other words, the findings suggest that students' awareness levels about the fact that these processes – including their potential undesirable effects – are possibly present in PA lead to placing greater value on peer assessment as a valuable learning activity. Furthermore, girls rated significantly higher on this factor than boys (Rotsaert et al., 2017a). It is important for teachers to be aware of the possible effect these perceptions can exert and, in a second stage, how to create an adequate classroom atmosphere in which fear of disapproval is not present. (3) *Psychological safety* refers to a shared belief by group members that there is a safe learning environment that enables different opinions to be perceived as opportunities rather than conflicts (Tapia & Fernández Heredia, 2008; Nicol, 2010; Yu & Sung, 2015). As with the factor *Fear of disapproval*, teachers' awareness about the importance of a safe classroom environment will be important to enable a positive classroom climate. (4) *Value congruency* refers to the importance of unanimity among assessors and assessees on both the goals and criteria of PA activity (Cheng & Tsai, 2012). Teachers' capacity to guarantee congruency about the used criteria will start from their awareness of the fact that this can significantly affect the outcomes of PA activity. (5) *Trust in oneself as an assessor* refers to the assessor's beliefs about his/her skills when assessing a peer (van Gennip, Segers, & Tillema, 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 has proved to be a significant predictor of students' perceptions of the educational value of PA (Rotsaert et al., 2017a). Again, teachers' capacity to build up trust in their own and others' evaluative capabilities will start from their level of awareness about this factor when implementing PA activities in their classroom.

Two factors are closely connected to the aforementioned interpersonal processes: the importance attributed to anonymity and perceived accuracy within PA. As regards 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. The idea behind this is that, as a result of the assessee's anonymity, the assessor focuses on the content, not whom they are assessing. Assessors' anonymity can help assessees focus on the feedback they receive, rather than on the person who gave the feedback. Vanderhoven, Raes, Montrieux, Rotsaert, and Schellens (2015) found that students have more positive attitudes toward PA when assessor anonymity is ensured. In this study it was also found that when assessors' names are not hidden from the teacher (while the assessor's anonymity was ensured amongst peers), this worked as a means of controlling for undesirable interpersonal effects such as friendship marking. Yu and Sung (2015) stated that anonymity might offer greater psychological safety for students, but at the same time, when anonymity is provided to the assessor, it might lead to misbehavior, such as positive marking towards friends. In his review, Panadero (2016) proposed that anonymity needs to be considered carefully when implementing PA because it might hinder formative uses of PA. Ideally, anonymity is approached as a temporary catalyst to create dialogic classroom environments in which students feel safe to participate (Rotsaert, Panadero, & Schellens, 2017). Importantly, Panadero and Brown (2017) revealed that the majority of their participant teachers used anonymous PA. The level of importance students attributed to

anonymity was found to be a significantly negative predictor of students' perception of the educational value of PA (Rotsaert et al., 2017a). Furthermore, in the same study, mean latent difference tests indicated that the importance attributed to anonymity was lower for students with three or more PA experiences than for those with fewer PA experiences (Rotsaert et al., 2017a).

Another crucial aspect related to interpersonal processes within PA is the concern about the (perceived) accuracy of all the actors in PA (Panadero, 2016). 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., 2017a).

1.3. 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 meso and micro levels, the Flemish government has emphasized the importance of autonomy and trust in the policymaking capacity of schools. As part of this autonomy, teachers and teacher councils are, as a rule, solely responsible for the majority of pupil's learning and classroom assessment. As a consequence, implementing formative assessment is the responsibility of individual teachers. Since 1998, the Flemish 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, the 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 a sample of Flemish teachers' current application of PA in their daily classroom practice.

1.4. Aim and research questions

The present study aim is to better understand the implementation of peer assessment done by Flemish teachers while exploring their perceptions about the social nature and educational value with a newly created instrument. Four research questions and goals are addressed:

RQ1– What is the current use and format of peer assessment of Flemish secondary teachers?

RQ2a – Validation of an instrument to measure teachers' level of awareness regarding students' perceptions of interpersonal processes within PA

RQ2b – Are teachers aware of students' perceptions of the interpersonal processes within PA?

RQ3 – What variables predict the educational value teachers attribute to PA?

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

2. Method

2.1. Participants

A total of 225 secondary school teachers participated in this study. Seventy-eight (34.66%) were teachers in grades 7–8 (henceforth *Level 1*), another 78 (34.66%) were in grades 9–10 (henceforth *Level 2*), and 69 (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.

2.2. Instrument

A self-report survey was designed with three blocks that was based on the instrument used by Rotsaert et al. (2017a) in a study on students' perceptions: (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. A pilot of this survey was conducted before it was administered. 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 level 2 (e.g. comprehension problems, length, etc.). After his input, the survey was then evaluated using a think-aloud procedure by two teachers from level 3. On average, they completed the questionnaire within 20 min.

The items in block C of this questionnaire used a six-point, positively packed rating scale in order to elicit more variance in responses (1 = strongly disagree, 2 = mostly disagree, 3 = slightly agree, 4 = moderately agree, 5 = mostly agree, and 6 = strongly agree) (Brown, 2004). The following definition of PA was presented to ensure that 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 in Supplementary material):

- 1 **PA activity description.** The first eight questions established an overview of the kind of PA activity teachers had organized, if any (block b).
- 2 **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).
- 3 **Importance of anonymity in PA.** Three items explored the degree to which teachers thought students found anonymity important in PA (block c).
- 4 **PA accuracy.** Two questions were used to establish whether teachers thought PA was an accurate assessment method (block c).
- 5 **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).

2.3. Procedure

The survey conductors were undergraduate students in educational

studies who were enrolled in a methodology course. In groups of four ($n = 31$), they were asked to obtain permission from eight teachers (equally spread over three school levels) to voluntarily conduct a survey in their classes. To facilitate this process, the students were offered a list of all secondary schools in Flanders from which they made a random selection.

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 manual with instructions (e.g. how to handle missing values), including a predefined Excel sheet to enter the data from the written surveys.

2.4. Analysis

To answer RQ1, we calculated the descriptive statistics from the data regarding teachers' previous PA organization in their classroom and the PA format. In line with Rotsaert et al. (2017a) and Panadero and Brown (2017), we report the data spread over three educational levels. Likewise, for RQ2b, after exploring the quality of the instrument and determining different factors through EFA and CFA analyses (RQ2a), descriptive statistics were calculated for the different scales. The goal of RQ3 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 to anonymity and teachers' perceived accuracy of PA. Therefore, we identified the relationships between several factors and explored their contribution to the perceived educational value of PA. For this, structural equation modeling (SEM) was used. SEM consists of a structural model representing the relationship between the latent variables and measurement models. These represent the relationship between the latent variables and manifest items, accounting for the measurement errors. Mplus software was used for these analyses.

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 an acceptable fit (Schreiber, Nora, Stage, Barlow, & King, 2006). 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 et al., 2006; Schweizer, 2010).

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 ΔCFI value higher than .01 to indicate a significant drop in fit, and Chen (2007) suggests using $\Delta RMSEA$ to test for evidence of invariance, the criteria for invariance were $\Delta CFI \leq .01$ and $\Delta RMSEA \leq .015$. If measurement invariance is achieved, researchers can accept that different groups of individuals interpret the items and their underlying constructs in similar ways.

3. Results

3.1. RQ1 – what is the current use and format of peer assessment of Flemish secondary teachers?

As shown in Table 2, 37.3% of participants at all secondary education levels had not implemented PA in their classrooms.

In Table 3 we report on the PA format implemented by the participants who had already organized PA in their classroom. Firstly, they

show a balanced use (yes/no/sometimes) of a PA score in the monthly report. Secondly, PA activities predominantly took place at the end of a series of lessons (on average 60.91%). Thirdly, 54.08% of all teachers indicated that they had not trained their students to implement PA.

Furthermore, the majority of teachers (70.37%) reported that they had not involved their students in defining the PA assessment criteria, which is seen as an important part of training in PA guidelines. The PA was mostly paper-based or conducted orally. Surprisingly, as several in-class (e.g. electronic response systems) and online (e.g. Mentimeter, Socrative) enablers have been developed in the last decade, technology-facilitated PA is almost totally absent (3.05%).

Table 4 provides details of teachers' responses regarding the implementation of anonymous PA activities. As can be seen, around half of the implemented PA activities provided anonymity for the assessor, but surprisingly, anonymity is hardly ever provided for the assessee. Finally, anonymity is never provided for the teacher, which is a positive finding as previous research indicated that the teacher's input is often used as a reference point when interpreting the assessment input by peers.

3.2. RQ2a – validation of an instrument to measure teachers' level of awareness regarding students' perceptions of interpersonal processes within PA

To answer research question 2a, first an exploratory principal axis factoring analysis was performed on a random 50% sample of the data. Following the recommendation of Hair, Black, Babin, Anderson, and Tatham (2006), all items with a loading of .50 or less were excluded from further analysis. Items were also removed where the factor loading differed by .25 or less on two factors. Such items were considered as having cross-loadings (Nunnally & 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 $\chi^2(55) = 509.73$, $p < .00$.

The first factor is "interpersonal processes 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 of the importance students attribute to guaranteeing anonymity for assessors and/or assessees. Teacher anonymity 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. The correlations between the three latent predictors are shown in Table 6.

Since our instrument taps several conceptually different constructs, each of which contributes 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

Table 2
Teachers' experience with PA.

Response category	Level 1 (n = 71)		Level 2 (n = 81)		Level 3 (n = 73)		Average (N = 225)	
	n	%	n	%	n	%	n	%
Never	26	36.61	36	44.44	22	30.14	84	37.33
Once	12	16.90	9	11.11	9	112.33	30	13.33
Twice	3	4.25	8	9.88	8	10.95	19	8.44
Thrice	9	12.68	6	7.41	5	6.85	20	8.90
> Thrice	21	29.58	22	27.16	29	339.73	72	32

Levels: Level 1 = Grade 7–8; Level 2 = Grade 9–10; Level 3 = Grade 11–12.

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.

3.3. RQ2b – are teachers aware of students' perceptions of the interpersonal processes within PA?

Table 7 presents the mean scores and standard deviation for each factor, ranging from a minimum score of 1 (strongly disagree) to a maximum of 6 (strongly agree). Results were slightly to moderately positive for the first three factors, meaning that teachers in this sample indicate that they are slightly to moderately aware of students' concerns related to interpersonal issues, anonymity, and accuracy within PA. Furthermore, they score highly positively for the attributed educational value of PA.

3.4. RQ3 – 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 at 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 Fig. 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 Fig. 1, only accuracy ($\beta = .565$, $std. error = .124$, $p < .001$) proved to have a significant impact on teachers' educational value of PA ($R^2 = .319$, $std. error = .141$).

4. Discussion

This study examined a sample of Flemish secondary school teachers' use of PA activities, their awareness of students' perceptions on the social nature of PA and how this influences the perceived PA educational value. The interpersonal nature of PA is an underexplored area and, additionally, research has tended to focus on students' perceptions with a few exceptions exploring those of teachers (Panadero, 2016). Importantly, a successful implementation of PA practices relies on the teacher's ability and motivation to prepare students adequately for PA (Harris & Brown, 2013; Panadero & Brown, 2017), hence the importance of exploring teachers' perceptions. A part of this preparation involves teachers' understanding of students' interpersonal concerns that may arise as a consequence of their participation in PA. Currently, however, little is known about teachers' awareness of the interpersonal dynamics within 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 of the 225 teachers had previously organized PA in their classroom, with a third at all three secondary levels having done so multiple times. It is worth noting that the reported PA organization is moderately higher than that found in previous studies. For example, in

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)		Average (n = 135)	
	n	%	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	53	39.27
No	12	29.27	22	48.89	17	34.70	51	37.77
Sometimes	11	26.83	7	15.56	13	26.53	31	22.96
Time: The PA activity took place								
...during a lesson	18	43.90	15	34.8	19	38.78	52	39.09
...at the end of a series of lessons	23	56.10	28	65.12	30	61.22	81	60.91
Training: Were students trained to implement PA?								
Yes	16	39.02	12	26.66	14	28.57	42	31.11
No	19	46.34	27	60.00	27	55.10	73	54.08
Sometimes	6	14.63	6	13.33	8	16.33	20	14.81
Involvement in defining PA criteria: Were students involved in defining the PA criteria?								
Yes	9	21.95	5	11.11	5	10.20	19	14.07
No	23	56.10	34	75.55	38	77.55	95	70.37
Sometimes	9	21.95	6	13.33	6	12.24	21	15.56
Via what tool was the PA conducted?								
Paper	19	46.34	18	40.00	29	59.18	66	50.38
Technology (laptop/computer...)	2	4.88	2	4.44	0	0	4	3.05
Oral	20	48.78	21	46.66	20	40.82	61	46.57

Table 4
Anonymity within PA activity.

Response category	Level 1 (n = 41)		Level 2 (n = 45)		Level 3 (n = 49)		Average (n = 135)	
	n	%	n	%	n	%	n	%
Anonymity for the assessor								
Yes	11	26.83	19	42.22	11	22.45	41	30.38
No	21	51.22	22	48.88	29	59.18	72	53.33
Sometimes	9	21.95	4	8.88	9	18.36	22	16.29
Anonymity for the assessee								
Yes	4	9.76	3	6.66	2	4.08	9	6.66
No	37	90.24	42	93.33	47	95.92	126	93.33
Sometimes	0	0	0	0	0	0	0	0
Anonymity for the teacher								
Yes	0	0	0	0	0	0	0	0
No	41	100	45	100	49	100	135	100
Sometimes	0	0	0	0	0	0	0	0

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 and Brown (2017). Our results confirm that PA is part of secondary Flemish teachers' armamentarium of innovative assessment methods. Although we cannot explore this in more detail with our data, the fact that the teachers reported low levels of both training and students' involvement in defining PA, and the fact that PA activities took place at the end of the series of lessons as a finalizing exercise, might indicate that there is still the potential to make PA practice even more formative (Fraile, Panadero, & Pardo, 2017; Sluijsmans, 2002).

In terms of anonymity, in contrast to the Spanish sample of teachers in Panadero and Brown's (2017) study, 53% (anonymity for the assessor) and 93% (anonymity for the assessee) of our group of teachers used nonanonymous modes of PA. This might indicate that the majority of our teachers are unaware about the fact that students might be resistant to participating in nonanonymous PA (cf. Rotsaert et al., 2017b) or underestimate the importance of a safe learning environment (cf. the lack of relationship that was found in RQ3 for both importance of anonymity and interpersonal factors within PA). The findings on anonymity are still inconclusive and highly dependent on the classroom climate (e.g. trust and mutual respect) and PA mode (e.g. face-to-face vs. online, synchronous vs. asynchronous), and future empirical studies are needed on this topic (Panadero, 2016; Rotsaert et al., 2017b). For example, a recent study focusing on students' perceptions shows that a temporary use of anonymous modes might be useful in offering space for students' associated need for practice in a safe environment (Rotsaert et al., 2017b). Nevertheless, there is still a tension to be resolved between anonymity and PA use: Future research needs to use more advanced models on how anonymity might affect different dimensions of peer assessment, instead of the simplistic on/off current approach (Panadero, 2016). With regard to RQ2a, EFA, CFA, and

Table 5
Results of the survey 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 fear 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 into, 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 consider it important that the identity of the assessee is hidden.	-.076	.963	.020	-.071
7	I think students consider 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/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.

Table 6
Correlation between latent predictors.

n = 244	Importance of anonymity	Accuracy
Interpersonal processes within PA	.403*	-.334*
Importance of anonymity		-.278**

* Correlations are significant at the .001 level.

** Correlation is significant at the .049 level.

reliability analyses were conducted so that four different factors could be identified: (1) 'interpersonal processes within PA,' (2) 'importance of anonymity', (3) 'accuracy', and (4) 'educational value of PA.' Goodness-of-fit estimates were calculated, all indicating acceptable fit. The first of the factors above refers 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 evaluates teachers' awareness of the importance students attribute to guaranteeing anonymity for assessors and/or assessees. The third factor measures teachers' perceptions of the accuracy of PA and whether they think students are capable of giving accurate judgments. The fourth factor refers to the extent to which teachers think PA is a valuable assessment method. This instrument was then used as a measure of teachers' awareness of students' perceptions of PA.

With respect to RQ2b, results show that our participants have a moderate awareness of students' concerns regarding the interpersonal processes in PA and of the importance students attribute to anonymity within PA. This indicates that the interpersonal challenges that were signaled in previous studies focusing on the students' perspective are not yet fully recognized by teachers at this point. Future qualitative research should explore whether this comes from a lack of knowledge or the fact that teachers do not find it influential on students' learning

Table 7
Survey factors' omega value, mean, and standard deviation.

	ω	M (SD)
Interpersonal processes 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)

Note. 1 = strongly disagree, 2 = mostly disagree, 3 = slightly agree, 4 = moderately agree, 5 = mostly agree, and 6 = strongly agree.

during PA activity. Furthermore, PA accuracy is slightly to moderately positively evaluated by teachers. Overall, the educational value of PA is highly positively appraised. Future research needs to explore what might be the influence of teachers' low reports of awareness in how they handle affective threats, such as discomfort and fear of disapproval and/or distrust derived from the implementation of PA (Higgins et al., 2001).

In general, it is for further research to explore how to raise teachers' attention 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 students' perceptions on influencing interpersonal processes, anonymity, perceived accuracy, and the value teachers attribute to PA (RQ3), SEM was used. Due to the lack of previous research on teachers' perceptions of the social nature of PA, our model is a first step in disentangling the complexity of this topic. The results lead to the conclusion that in our sample 31.9% of the variance in the Educational value of PA can be solely explained by teachers' perceived accuracy of PA. It is logical that our sample considers PA accuracy to be crucial as it has been of major concern for both

researchers and practitioners (Panadero et al., 2013; Topping, 2003). Our two accuracy items (Students are capable of giving an accurate judgment to each other, and Evaluation through PA is accurate) show that teachers really value the fact that the feedback given through PA should be accurate for it to have educational value. In Panadero, Jonsson, and Strijbos's review (2016) it was emphasized that in terms of scoring accuracy (i.e. the degree of closeness between the peer score and the teacher score) it might be more valuable to achieve PA content accuracy (i.e. the degree of closeness in the qualitative feedback between a peer and a teacher). In this study, we cannot figure out whether our teachers value both accuracies equally, but it seems logical that they consider the accuracy of PA to be a crucial component of the educational value.

Additionally, comparing our results with those of Panadero and Brown (2017), they found as predictors of teachers' PA experience and frequency of use: positive experience with PA (.61), educational PA advantages (.53), previous use of PA (.51), willingness to use PA in grading (.29), and belief in student participation in assessment (.27). Importantly, those authors explored primary, secondary, and higher education teachers and found strong differences in how PA was implemented in the different levels. Therefore, from the present study data, it can be concluded that for secondary teachers to consider PA to have an educational value, PA accuracy is crucial.

However, the fact that accuracy was the only identifiable predicting factor in our study raises concern. In other words, 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 an important predictor of students' perception of the educational value of PA (Rotsaert et al., 2017a), the discrepancy between students' perceptions of interpersonal variables and related aspects and teachers' awareness level regarding these students' perceptions should raise concerns. This 'misalignment' has important consequences for teacher training and professional development in assessment literacy.

4.1. Limitations

One important limitation of this study is its self-reported nature. As this study only uses a measure of teachers' awareness levels regarding

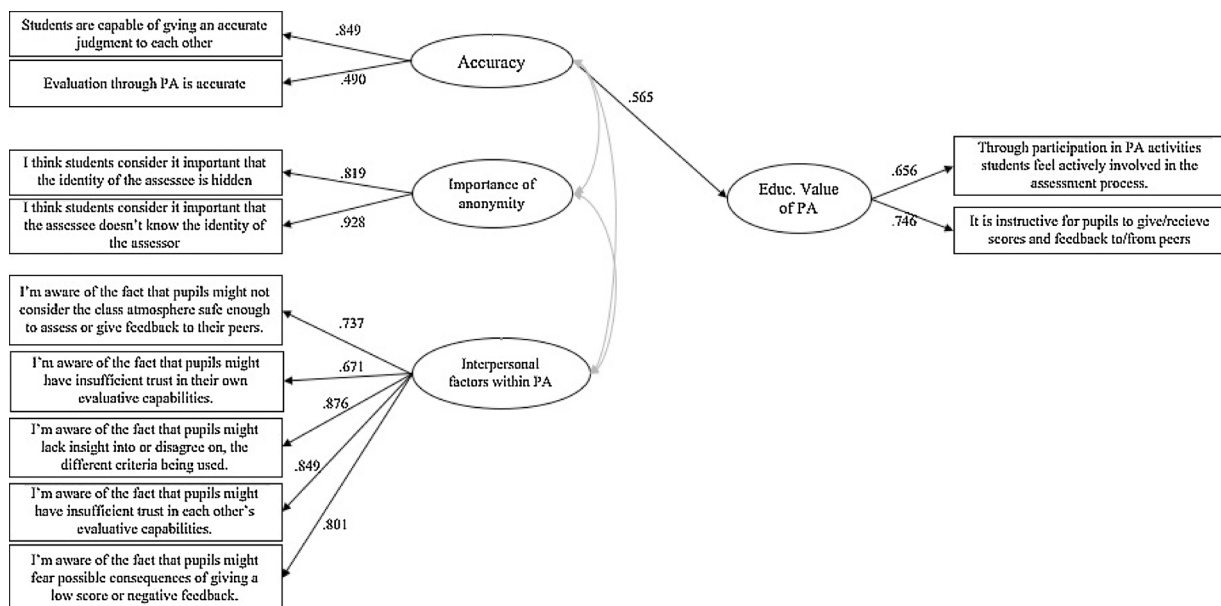


Fig. 1. Educational value of PA baseline SEM model with standardized loading.

interpersonal processes, future studies should 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 self-reported data (Desimone, 2009). However, throughout the process of survey development and administration, several steps were taken to reduce social desirability bias (cf. positively packed rating scales) and ensure confidentiality for respondents.

Another limitation might be that some important aspects of the latent constructs were not adequately highlighted by the questions in our instrument. For example, our study involved complex psychological constructs, and the model included a limited number of items for measuring each one. However, 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 other diverse measures for these same constructs in future studies, such as the use of qualitative data (e.g. video analyses of the evolution in a prosocial classroom atmosphere or specific teacher moderator actions). In sum, based on the incentives that were given by Panadero (2016), we have explored the first model to look at the role of interpersonal processes in PA from a teacher's point of view. This model should be seen as a starting point to be further developed. We as authors are convinced that future mixed-methods design studies will be able to help us grasp better the interrelatedness of these processes.

4.2. Implications and future lines of research

Based on the finding that teachers' perceived accuracy of PA is crucial for the educational value they place on PA, researchers need to explore teachers' understanding of PA accuracy and whether it includes content accuracy or just scoring accuracy (Panadero, 2016). More importantly, the pre- and in-service teacher training programs should aim to work on how to implement PA in ways that increase students' PA accuracy. This type of training should be based on assessment for learning methods in which the distinction between assessment and instruction has been blurred because good assessment is a reflection of a good instructional setting (Brookhart & Nitko, 2013). In this view, (peer) assessment is conceptualized 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 implies that PA should no longer be seen as an episodic mechanism but rather as an instructional approach that calls for a central position in the curriculum (Panadero & Brown, 2017; Panadero, 2016; Rowe, 2017). 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 high-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 et al., 2016, p. 10). In doing this, involvement in PA has the potential to promote prosocial behavior, ease the effect of negative emotions, and contribute to a sense of belonging (Panadero, 2016; Rowe, 2017).

5. Conclusions

This study adds to our understanding of teachers' use of PA and opens up a new avenue 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' awareness and mastery of these interpersonal processes. More specifically, our study points out that teachers' perceived accuracy is a major predictor of their belief in the educational value of PA.

Funding

The first author's research was funded by Ghent University BOF fund number BOF13/24J/115.

The second author's research was funded by the Spanish Ramón y Cajal program number RYC-2013-13469.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.stueduc.2018.07.001>.

References

- Ajzen, I. (2005). *Attitudes, personality, and behavior* (2nd ed.). New York: Open University Press.
- 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. <https://doi.org/10.1111/jcal.12091>.
- 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*.
- Brookhart, S. M., & Nitko, A. J. (2013). *Educational assessment of students: Pearson new international edition*. Pearson Higher Ed.
- 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. <https://doi.org/10.2466/pr0.94.3.1015-1024>.
- Brown, G. T. L., & Harris, L. R. (2016). *Handbook of human and social conditions in assessment*. New York, NY: Routledge.
- 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. <https://doi.org/10.1080/0969594X.2014.996523>.
- Carless, D. (2013). *Trust and its role in facilitating dialogic feedback*. *Feedback in higher and professional education* 90–103. <https://doi.org/10.1080/02602931003632381>.
- Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment and Evaluation in Higher Education*, 35(5), 551–564. <https://doi.org/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.
- 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.
- 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. <https://doi.org/10.1080/07294360701494328>.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <https://doi.org/10.3102/0013189X08331140>.
- Dochy, F., Segers, M., & Slujsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24, 331–350. <https://doi.org/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. <https://doi.org/10.3102/00346543070003287>.
- Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: Useful or waste of time? The effects of establishing assessment criteria with students on self-regulation, self-efficacy and performance. *Studies in Educational Evaluation*, 53, 69–76. <https://doi.org/10.1016/j.stueduc.2017.03.003>.
- Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). *Multivariate data analysis* (6th ed.). Upper saddle River, NJ: 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. <https://doi.org/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. <https://doi.org/10.1080/13562510120045230>.
- Lynch, D. H., & Golen, S. (1992). Peer evaluation of writing in business communication classes. *The Journal of Education for Business*, 68(1), 44–48.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.J.: Erlbaum Associates.
- Murdoch, 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.
- Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment and Evaluation in Higher Education*, 35(5), 501–517. <https://doi.org/10.1080/02602931003786559>.
- Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment and Evaluation in Higher Education*, 39(1), 102–122. <https://doi.org/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. <https://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. Assessment for learning: Meeting the challenge of implementation* 311–326. <https://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. <https://doi.org/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. <https://doi.org/10.1080/03075079.2013.823930>.
- 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. <https://doi.org/10.1080/02602938.2015.1008982>.
- Rotsaert, T., Panadero, E., Estrada, E., & Schellens, T. (2017a). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. In P. Van Petegem (Vol. Ed.), *Studies in Educational Evaluation: Vol. 53*, (pp. 29–40).
- Rotsaert, T., Panadero, E., & Schellens, T. (2017b). Anonymity as an instructional scaffold in peer assessment: its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. In A. Wichmann, & J.-W. Strijbos (Eds.), *European Journal of Psychology of Education*.
- Rowe, A. D. (2017). *Feelings about feedback: The role of emotions in assessment for learning*. Singapore: Springer 159–172. https://doi.org/10.1007/978-981-10-3045-1_11.
- Rubie-Davies, C. M., Flint, A., & McDonald, L. G. (2012). Teacher beliefs, teacher characteristics, and school contextual factors: What are the relationships? *British Journal of Educational Psychology*, 82, 270–288. <https://doi.org/10.1111/j.2044-8279.2011.02025.x>.
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and 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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1007/s11336-008-9101-0>.
- Sluijsmans, D. (2002). *Student involvement in assessment: The training of peer assessment skills*. Heerlen: Open Universiteit.
- Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, 80, 217–222. https://doi.org/10.1207/S15327752JPA8003_01.
- Tapia, J. A., & Fernández Heredia, B. (2008). Development and initial validation of the classroom motivational climate questionnaire (CMCQ). *Psicothema*, 20(4).
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249. <https://doi.org/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 (Vol. Eds.), *Optimising new modes of assessment: In search of qualities and standards: Vol. 1*, (pp. 55–87). Netherlands: Springer.
- Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, 20(4), 339–343. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1016/j.compedu.2014.10.001>.
- Vlaams Ministerie van Onderwijs en Vorming – Onderwijsinspectie (2014). *Onderwijs Spiegel*. Retrieved from http://www.onderwijsinspectie.be/sites/default/files/atoms/files/Onderwijs Spiegel_2014.pdf.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>.
- Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, 18(3), 1–13. <https://doi.org/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. <https://doi.org/10.1080/10494820.2015.1041405>.