Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions

Ernesto Panadero [1]

Author Note

[1] Departamento de Psicología Evolutiva y de la Educación. Universidad Autónoma de Madrid (Spain).

Recommended citation

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 Social and Human Conditions in Assessment* (pp. 247-266). New York: Routledge.

Correspondence concerning this article should be addressed to: Ernesto Panadero. Despacho 109, Departamento de Psicología Evolutiva y de la Educación, Facultad de Psicología, Universidad Autónoma de Madrid, Cantoblanco (Madrid), 28049 Spain. E-mail: ernesto.panadero@uam.es. Phone (+34) 914973553.

Acknowledgements: Research funded by the Spanish Ministry (Ministerio de Economía y Competitividad) via Ramón y Cajal funding granted to the author (File id. RYC-2013-13469). Special thanks to Jan-Willem Strijbos for being always on the "other side" and helping me reflect about the content from this chapter. Also thanks to the editors, Gavin Brown and Lois Harris for their great job editing the chapter, and to Keith Topping and Jeltsen Peeters who helped providing inaccessible manuscripts.

"With great power there must also come great responsibility".

(Uncle Ben to Peter Parker in Lee & Ditko, 1962)

1. **Introduction**

Peer assessment occurs when people of equal status assess each other's work; most commonly in education, the peer is a classmate or a student from the same institution. Assessing a peer activates a significant number of cognitive, motivational, and emotional processes and has potential to enhance the assessor's and the assessee's learning (van Gennip, Segers, & Tillema, 2009). Mainly for this reason, formative assessment / assessment for learning (FA/AfL) advocates, including myself, consider peer assessment, along with self-assessment, to be central aspects of FA/AfL practices (Panadero & Alonso-Tapia, 2013; Reinholz, 2015). Additionally, peer assessment gives students power in the assessment process; that is, they can express their perspective and share responsibility (Dixon, Hawe, & Parr, 2011). This involves multiple social and human factors that need to be taken into account because peer assessment does not happen in a vacuum; rather it produces thoughts, actions, and emotions as a consequence of the interaction of assessees and assessors.

While there have been many studies on peer assessment, the social and human factors have been relatively overlooked and attention to these dimensions is relatively recent (Strijbos et al., 2009; van Gennip et al., 2009). Two reasons might explain this growing interest: (1) increased implementation of PA as a result of Black and Wiliam's (1998) seminal review on formative assessment and (2) van Gennip et al.'s (2009) review of interpersonal and social issues in PA. While the van Gennip et al. paper reviewed much research on PA and learning outcomes, it was a seminal synthesis of the four articles available at the time that examined interpersonal variables; after this, there was an increase in the number of studies on the topic. This chapter will first examine

definitions of peer assessment (PA), followed by a review of the main findings from reliability research into "traditional" PA (i.e., giving an evaluative score, grade, or rating, with little or no clarification or justification for the evaluation). Then, research that explores social aspects of PA and "new approaches" to PA (i.e., giving rationale, explanation, or descriptions that inform improvements) will be reviewed. In this research, if peer scoring accuracy is explored it is always in interaction with other aspects (e.g. Panadero, Romero & Strijbos, 2013) amplifying our understanding of the phenomena. Because this new PA research has a shorter tradition, I will conclude with specific recommendations for future research.

2.1. Definition of peer assessment and learning benefits

Definitions of peer assessment range from the process whereby groups of individuals rate their peers (Dochy, Segers, & Sluijsmans, 1999; Falchikov, 1995) to a consideration of "the amount, level, value, worth, quality or success of the products or outcomes of learning of peers of similar status" (Topping, 1998, p. 250). Others would consider it to be a form of collaborative learning (Kollar & Fischer, 2010; Strijbos, Ochoa, Sluijsmans, Segers, & Tillema, 2009). PA can range from asking the students to give a grade to a peer's work (i.e., peer grading/marking) (Falchikov, 1995) to expecting them to give and/or discuss specific feedback (Strijbos et al., 2009). Different PA practices seem to lead to differences in the depth of learning and the use of self-regulated learning strategies (van Gennip, 2012a). The idea that there are different types of PA is not new (Topping, 1998, 2012; van Gennip et al., 2009; Van Zundert, Sluijsmans, & Van Merriënboer, 2010). The effects of different PA forms have been studied; for example, the advantages of using a rubric to support PA over an open-ended format of PA (i.e., unguided asking the students to PA) (Panadero, Romero & Strijbos,

2013) and the benefits of guided PA where students discuss the role of the assessor and how to grade peers' work (van Gennip, 2012a) have both been established.

A crucial argument for the use of PA is that it enhances both learning and performance in two distinct ways (Corgnet, 2012; Topping, 1998; van Gennip et al., 2009; van Zundert et al, 2010). First, the assessee whose work is assessed receives direct feedback on how to improve. Second, the assessor, by evaluating a peer's work, becomes more aware of his or her own strengths and weaknesses. In other words, peer assessment enhances self-assessment capability (Dochy et al., 1999; Reinholz, 2015) and students' self-regulated learning, with peers acting as co-regulators of their assessee peers (Panadero, Jonsson & Strijbos, 2016). Hence, PA is an opportunity for students to learn more. Overall, most students have positive attitudes towards PA and report having learning gains (Bryant & Carless, 2010; Orsmond, Merry, & Reiling, 1996; van Zundert et al., 2010), even though those gains are not always substantiated in empirical studies (Li & Steckelberg, 2004).

2.2. Reliability research into peer assessment

Traditionally, there has been considerable research into the reliability and/or validity and/or accuracy of peer scores (e.g. Falchikov & Goldfinch, 2000). Distinctions between these three constructs are relatively subtle (Panadero et al., 2013). Reliability generally focuses on the consistency of student peer marks with teacher or tutor marking. Validity has to do with the appropriateness of peer comments, corrections, or grades. Accuracy has to do with the truthfulness or veridicality of peer assessment. In other words, good PA might be that which (a) corresponds with teacher evaluations, (b) is appropriate to the work being evaluated, and (c) aligns with the reality of the work under consideration. More detailed discussions of these distinctions are available in Panadero, et al. (2013) and to a lesser extent in Topping (2003, 2012). Peers can

produce assessments in a manner consistent with their teachers (Topping, 2012) as demonstrated across different narrative reviews (Topping, 1998, 2003, 2012; van Zundert et al., 2010) and a meta-analysis (Falchikov & Goldfinch, 2000). The reliability of PA is greater when a number of pedagogical conditions are present, such as involving students in discussion about the criteria, avoiding a large numbers of peers in a peer assessment group, considering the level of expertise of the students, practicing how to peer assess over time, and receiving teachers' feedback about the PA conducted (Falchikov & Goldfinch, 2000; Strijbos et al., 2009; Topping, 2003).

However, the focus of PA research has been dominated by one practice; that is, "peer rating … often in combination with open comments" (Strijbos et al., 2009, p. 388). Additionally, the focus of PA research has been on summative purposes (e.g., aspects of grading); Topping (1998) reported that in 31 higher education studies that the focus was "on scores and grades awarded by peers rather than through a focus on more open ended formative feedback" (p. 257). It may be that research into the reliability of PA continues, despite its robust establishment, because, although researchers and instructors would like to include PA as a legitimate part of student summative evaluations to improve assessment efficiency and validity, they still feel uncomfortable doing so given assessment is traditionally viewed as a teacher's responsibility (Magin, 2001). Potentially, if students are accurate peer assessors, then teachers could use PA to save time (Sadler & Good, 2006); however, since teachers have to devote time to teach students how to do PA and give them opportunities to practice (Strijbos et al., 2009; Topping, 1998), it remains unclear if instructor time would actually be saved via these processes. Additionally, hardly ever is there any consideration of how PA accuracy might directly affect the learning of the assessor or assessee (Topping, 1998) which is a crucial aspect that this type of PA research has yet to explore.

Moreover, students and teachers seem to not yet feel comfortable in allowing a peer or student score count towards the final grade in a course (Panadero & Brown, 2015; Wen & Tsai, 2006). Therefore, we need to clarify the social and interpersonal variables that play a role in PA.

3. Interpersonal research into PA: when social aspects matters and formative uses of PA prevail

Because PA can be conceptualized as a collaborative learning activity, it is clear that the social aspects of PA need to be considered, as varied PA practices will elicit different interpersonal variables (van Gennip, 2012a). Attention to social factors is needed because well implemented PA should decrease negative social problems (e.g., negative situations like tension among peers because of peer scoring) (Topping, 2012).

3.1 Social actors in PA: Students and teachers

3.1.1 Students. The roles of assessor and assessee are not identical and have different psychological and interpersonal consequences (Panadero et al., 2013; van Zundert et al., 2010). As the assessor, the student must pass some sort of judgment on a peer and as an assessee the student must experience, and hopefully learn from, that judgment. Research that has differentiated between these contrasting roles is scarce (van Zundert et al., 2010).

Panadero et al. (2013) explored correlations between comfort as assessor and perceived fairness as assessee, finding small correlations, supporting the contention that students distinguish between the two roles. Further, the same study found students in PA groups gave moderately positive ratings of comfort when acting as assessor, and more negative comfort attitudes as assessee. This reluctance to being assessed by peers has been found elsewhere where students have indicated a preference for teacher judgments (Bryant & Carless, 2010; Gao, 2009; Harris & Brown, 2013; Peterson &

Irving, 2008). A significant problem seems to be authority; that is, students do not perceive their peers as an authoritative source unlike more traditional sources (e.g., instructors or books) (Strijbos et al., 2009). In sum, the main actors of PA, the students, have some concerns, not with the process of PA itself or how much they can learn from it, but with their roles as assessors and assessees.

3.1.2 Teachers. Overall, teachers are positive about PA, reporting that it enhances learning and self-regulated learning strategies (Bryant & Carless, 2010; Harris & Brown, 2013; Lindblom-Ylänne, Pihlajamäki, & Kotkas, 2006; Panadero & Brown, 2015; Topping, 2012). Additionally, pre-service teachers appear to hold similarly positive views that PA can enhance learning and the use of learning strategies (Karaca, 2009; Koc, 2011; Wen & Tsai, 2006). However, recent research has shown that teachers consider PA ineffective in reducing their workload. For example, Panadero and Brown (2015) found that 95% of 1510 Spanish teachers reported that PA did not save them any time. Additionally, one-third to half the teachers indicated problems with PA having to do with reliability, student mistrust of peer scores, and the creation of social problems in classroom groups.

Teachers may also need to consider parents' perceptions of PA. The 2001 U.S. Supreme Court Case of Falvo v. Owasso School System (Sadler & Good, 2006) examined the appropriateness of peer assessment; in this situation, a sixth-grade boy with learning disabilities was teased after receiving low scores on a peer-graded quiz. This case resulted in a federal ruling that prohibited peer-grading in six states of the USA. Sadler and Good (2006) noted that a case about the appropriateness of disclosing grades to classmates via peer grading turned into a case against peer-grading in itself. Since peer-assessment involves children, the opinions and attitudes of their parents also

matter for successful implementation of PA, an issue explored more by Buckendahl (this volume).

4. Human and social variables in peer assessment: empirical evidence

4.1. Review criteria

In searching the literature, the following criteria were used: (1) empirical evidence of the effect of PA on social or human factors or vice versa had to be available; (2) the study had to be to published and peer-reviewed, including doctoral dissertations; and (3) articles had to be available in English. Articles related to clinical psychology applications were excluded. Searches in PsycINFO were conducted without and with year restriction (>2009) for the following combinations of key terms: PA + 'motivation'; + 'emotion'; + 'interpersonal'; + 'psychological safety'; + 'friendship'; + 'value diversity'; + 'value congruence' in abstracts. While anonymity is a crucial variable in PA (e.g., Vanderhoven et al., 2015), it has not been included in the search because few studies exploring its effect on interpersonal and human variables have been conducted. The searches identified a total of 191 abstracts which were read and 16 relevant articles were identified. Another 10 relevant papers were identified from the reference lists of papers already included in the review. The basic content of the 26 selected articles is captured in Table 1 and is used to create a narrative review following Dochy's (2006) guidelines. A meta-analysis could not be conducted due to the lack of consistent definition of PA and the high number of descriptive studies.

Table 1 lists the articles by the author's name in chronological order of publication. The Sample column reports information regarding the study sample as provided by the authors (i.e., number of students, mean age, gender distribution, educational level, country and subject of the PA implementation). The research method and data collection column captures the study design and instrumentation information as

supplied by the authors. With regards to the classification of the research method used, the taxonomy by Montero and León (2007) was used. Four types of research methods have been used: (a) experimental with random assignation of individual participants to the experimental conditions (Experimental design with Random Group, ERG), (b) experimental with random assignation of classroom groups to the experimental conditions (Experimental design with Random Intact Groups, ERIG), (c) implementation of PA in an educational setting to explore the effects with no control group (Descriptive design with Structured Observation, DSO), and (d) a qualitative approach design using case study (Multiple Case Study, MCS). The Variables of Interest column presents the focus of the study as relevant to this chapter. The Type of PA column describes the relevant characteristics of the PA implementation, including the task for PA, whether PA was anonymous or not, the use of scoring tools, and whether feedback was face to face or not, and so on. It is important to point out that differences in the quality and quantity of information provided in the table are attributable primarily to the differences in the level of detail in the original manuscripts study (e.g. if the details of the PA implementation were missing, that cell would have less information).

Table 1. *Summary of studies on PA and interpersonal variables or human factors* (chronological order of publication)

#	Study	Sample	Research Method & Data Collection	Variables of interest	Type of PA
1	Falchikov, 1995	*N*=13, 12 female; age *M* = 20 years, 11 months. Students in developmental psychology course for degree in biological science. Scotland.	DSO. Self-reported data via an instrument (named self-evaluation sheet) with 7 questions (3 open ones and 4 continuum choices e.g. fair vs. unfair)	Fairness and friendship	Short oral presentations were evaluated by classmates. Assessors were anonymous. Students were given list of 4 criteria plus a form to identify 'best feature' and 'weaknesses'. Scores were given in a 20 points scale and they counted towards the final grade.
2	Hunter & Russ, 1996	Second and last year university music students. Ireland.	DSO Data are "comments from the participants".	Fairness and friendship	Musical performances were PA, first by same year students, later by more senior students. PA was scaffolded with report outlines and criteria. Scoring agreement was checked.
3	Stanier, 1997	*N*=36; university students in geography courses. United Kingdom	DSO. It seems to be self-reported data via an evaluation questionnaire with 6 questions but also there is more data although source is not reported.	Motivation and students' perceptions for PA.	Group work was peer assessed. Assessors used a discussed set of criteria.
4	Topping, Smith, Swanson & Elliot, 2000	*N*=12; 10 female; Age *M* = 31. Educational psychology post-	DSO. Data are analysis of the students PA feedback forms	Discomfort	A written academic report was PA. Reciprocal PA by pairing the students was implemented with

		graduate students. Scotland.	and questionnaire with 24 question including "choose an option" (e.g. Y/N) and open questions		face to face interactions and detailed feedback. Assessors filled out a report also filled by the staff and later exchanged with the assessee.
5	Lin, Liu & Yuan, 2002	*N*=57; 88% male, senior students vocational school and *N*=41 (85% male) university students. Taiwan.	DSO. Comparison between secondary and higher education students' attitudes towards PA. Self-reported, questionnaire with 11 questions 5-point Likert items.	Trust in other as assessor.	Different PA procedures for secondary and university students. Random assignment of assessors that used given criteria to assess. Only the university students could make changes after online peer feedback.
6	Sluijsmans, Brand-Gruwel & van Merriënboer, 2002	*N*=93; 19 male; age *M*=20.7. Pre-service teachers. The Netherlands.	ERIG. Self-reported data: questionnaire 72 items 5-points Likert scale and forms.	Trust in self as assessor.	PA of videotaped lessons created by groups. Each student assessed 3 groups and gave 'qualitative PA'.
7	Li & Steckelberg, 2004	*N*=48. Students in Instructional Technology course. USA	ERG. Self-reported survey with 11 items, 5-point Likert.	Trust in other as assessor.	Webquests were assessed. Online PA with randomly chosen assessors that used a given rubric. Assessees could, after PA, make changes to their work.
8	Wen & Tsai, 2006	*N*=280; Male 58%.pre-service teachers. Taiwan	DSO. Self-reported survey on perceptions of PA vs. online PA (34 5points-Likert items)	Fairness	The study was exploring perceptions on PA from previous learning experience. There was no PA implementation within the study.
9	McMahon, 2009	*N*=5. University students enrolled in a Bachelor degree or a	MCS. Self-reported data via 'structured critical reflection'	Motivation and confidence as assessor.	PA was organized in recursive iterations within the five participants meetings. PA was

		Certificate of Higher Education. Ireland.			non-anonymous with detailed feedback in one to one and group instances.
10	Johnson & Winterbottom, 2010	*N*=28 girls from a rural comprehensive school aged 11-18. UK	DSO. Self-report data (questionnaires: PALS & IMI), interviews, classroom observations and video-records of PA events. Students grouped via prior attainment	Motivation (goal orientation)	Face to face PA opportunities were given.
11	McConlogue, 2010	*N*=82 engineering students. UK	DSO. Self-report data via a questionnaires and focus group. PA implemented with two different samples (i.e., Level 5 and Level 7 of the curriculum).	Fairness	Level 5 assessed 4 to 5 anonymised peer's coursework assignments, each assignment evaluated by 4 to 5 assessors. Assessees received their work with all grades and feedback. In the Level 7 group, the PA was identical with the addition of prior discussion of assessment criteria.
12	van Gennip, Gijbels, Segers & Tillema, 2010	*N*=118 employees from 4 organizations. The Netherlands.	DSO. Self-report data (questionnaires).	Trust, psychological safety, value diversity, interdependence, and transparency	Each participant was assessed on their work performance via questionnaires by 3 or 4 persons (work colleague, manager, inferior or customer) in a 360º feedback process.
13	van Gennip, Segers & Tillema, 2010	*N*=62 male secondary-vocational education students	ERG. Control and PA condition. 17 groups of 3 to 5 students in a project-based	Trust, psychological safety, value diversity, interdependence, and	Experimental condition received PA training. Each student evaluated the other groups work using a form including agreed

		aged 16-19. The Netherlands.	course. Only experimental groups received PA training. Self-reported data using validated questionnaires.	predictors of perceived learning.	criteria as a classroom. Control group was evaluated by teacher.
14	Willey & Gardner, 2010	*N*=89 engineering students. Australia	DSO. Self-report online survey had mixture of free response and 4-point rating items	Fairness and comfort	SPARK^{PLUS} software for self-assessment and PA. Software provided anonymous written feedback and scores.
15	Carvalho, 2012	*N*=120; 73% age 21-25 years. Business higher education students, 7 different cohorts, in teams of 5-8 members. Portugal.	DSO. Self-report survey data with 4liker-point items.	Friendship-marking, fairness and comfort.	PA among teammates working on a collaborative learning activity.
16	Cheng & Tsai, 2012	*N*=23. First year university students. Age *M*=18.74, 8 males. Taiwan	DSO. Self-reported data (interview). Note: students were also asked about their conceptions and learning approaches to explore how this interacted with the PA variables studied.	Trust in the self and the other, psychological safety, interdependence and value diversity	Online anonymous PA with grading (100 point scale) and "comment making" (feedback). Each student assessed 5 classmates. PA was done in five different tasks, so each student received 25 PA evaluations during the semester.
17	Corgnet, 2012	*N*= 66. University students. Spain.	ERG. 3 conditions: Asymmetric Peer Evaluations (APE); Symmetric Peer Evaluations (SPE); Equal Split Conditions (EQS). Self-report (Questionnaire).	Friendship, motivation	The peers evaluated members of same work group using a questionnaire.

18	Hou & Cheng, 2012	*N*=65. University students in multimedia production working in groups of 3.Taiwan.	DSO. Self-report (questionnaire)	Emotional state	Online forum commentaries on videos created and posted online by small groups of two or three students.
19	van Gennip, 2012b	*N*=106, males; Ages 15-18. Vocational school students in project-based course. The Netherlands.	ERIG. 3 conditions: control, PA, and PA+ (enhanced). Self-report (questionnaire).	Psychological safety and value congruency	PA conditions received PA training and PA+ condition received extra training on the role of assessors. Group work final product was assessed by the rest of classmates using a form with 11 criteria.
20	Harris & Brown, 2013	Three teachers and their students from primary, intermediate and secondary education. New Zealand.	MCS. Classroom observation (narrative field notes), further analysis of the video recordings, students' work samples, focus groups interviews and individual teacher interviews.	Social factors in students' perceptions on PA (e.g., friendship bias)	Researchers observed natural instances of PA in the classroom including rubric-guided essay marking.
21	Kim & Ryu, 2013	N=122 Pre-service teachers aged M=21.57 (*SD*=1.4). Korea.	ERG. 3 conditions: enhanced web-based PA, traditional PA, and traditional self-assessment	Metacognition, performance and motivation (confidence and satisfaction)	Traditional PA condition = rubric based scoring of non-anonymous peer work. Enhanced PA = online use of WFPAS software.
22	Panadero, Romero & Strijbos, 2013	*N*=209; 87% female pre-service teachers in "Learning and development II	ERIG. Two conditions created from 4 classroom groups assigned randomly to: control PA and rubric-PA.	Friendship construct validity bias, perceived comfort as PA assessor, and	Control PA = gave a grade and feedback to the work from a non-anonymous peer. Rubric PA =

		course" aged *M*=22.17 (*SD*=3.92). Spain	Self-report data (questionnaires and open questions).	perceived fairness as assessee.	rubric with different criteria guided grading and feedback.
23	Cheng et al., 2014	*N*=65; 58 female, University students in Advanced communication technology workshop aged 18-19. Taiwan.	DSO. Students worked in 21 groups of 2, 3 or 4 members. Self-report data on emotions based on the PA received (using the Self-assessment manikin).	Emotions: emotional responses to PA comments	Peers commented anonymously online to group projects on YouTube. Assessees could decide to reply to those comments or not.
24	Hwang et al., 2014	*N*=167; age 12. Elementary students in 10-week natural science course. Taiwan.	ERIG. Control and experimental conditions, three classroom groups assigned to each. Self-report data via: two scales, one questionnaire and one open-ended questionnaire.	Motivation.	PA condition = random assignment of work marked according to guidelines. Control group = no PA, instructor assessed work.
25	Vanderhoven, Raes, Montrieux, Rotsaert, & Schellens, 2015	*N*=69; 72% girls; age range 15-16. Grade 9 and 10 secondary students in groups of 2 or 3. Belgium.	ERIG. Two conditions: Traditional non-anonymous PA and Experimental with anonymous computer assisted PA. Mixed-method data: self-reported questionnaire, observation of the logbook, semi-structured interviews with teacher and students, and open question at the end of the questionnaire.	Anonymity in relation to social effects (peer pressure and fear of disapproval), discomfort with PA and emotions towards PA.	Small group presentations after PA training. Traditional non-anonymous PA = public "raising score cards" to evaluate presentation. Experimental, anonymous = computerized system to evaluate peer presentations.

26	Wilson et al., 2015	*N*=228. University students. Australia.	DSO. Self-report: survey, interview to selected students and 11 teachers	Fairness and workload (temporality for students).	Anonymous PA using a massive online tool serving 2793 students from different courses.

Note. PA=peer assessment or peer assessed; ERG= Experimental design with Random Group; ERIG=Experimental design with Random Intact Groups; DSO=Descriptive design with Structured Observation; MCS=Multiple Case Study; PALS= Patterns of Adaptive Learning Scales; IMI= Intrinsic Motivational Inventory; WFPAS= Web-based Formative Peer Assessment System.

Ten themes were identified in the selected studies:

(1) motivation, having to do with the reasons or goals students have for performing PA and the effects of PA on motivation;

(2) emotion, focused on the feelings connected to PA and its implementation;

(3) social factors, related to aspects of social connections, such as enhancing peer scores to be more liked by classmates;

(4) friendship, the potentially biasing effect of prior relations on PA (Note: while this is also a social factor, a number of studies exploring it have made it the most salient factor, so it was decided to treat it as its own theme);

(5) trust in the self as assessor, focusing on belief that one can be a good assessor;

(6) trust in the other as assessor, related to having confidence in a peer who is capable of performing PA;

(7) fairness/(dis)comfort, referring to beliefs about the appropriateness of PA and the student's level of comfort within assessor and/or assessee roles;

(8) psychological safety, focusing on the extent to which students feel safe to give sincere feedback as an assessor and do not fear inappropriate negative feedback as an assessee;

(9) value diversity/congruence, referring to the level of goal similarity between assessor and assessee; and

(10) interdependence, examining the degree to which assessors and assessees are mutually dependent on each other.

However, because of similarity between these themes, the empirical evidence is aggregated in three major themes. Intra-individual motivations, emotions, perceptions of fairness, and discomfort are grouped into *intra-individual factors in PA*, while, inter-

individual social factors, friendship, and safety are grouped into *inter-personal aspects of PA*. The third group consists of cognitive issues to do with the confidence or trust in the competence of the peer as assessor, the self as an assessor, group commitment to values, and sense of task interdependence, referred to as *cognitive aspects of PA*.

4.2. Intra-individual factors in PA

PA is usually conceptualized as an activity that enhances students' motivation as they take more control of their learning (e.g., Topping, 1998). However, as students might also feel powerful negative emotions (e.g., embarrassment) because of classmate feedback, these negative experiences have to be considered in examining students' motivations around PA. Of the six studies exploring motivation, four focused on a general concept of motivation; that is, if students were satisfied and willing to perform PA in the future (Hwang et al., 2014; Kim & Ryu; 2013, McMahon, 2009; Stanier, 1997), all of which reported students liking PA. Only two studies explored more specific aspects of motivation. Johnson and Winterbottom (2010) found that after implementing PA (and also self-assessment), students self-reported a decrease in learning goals, although the qualitative classroom observations suggested they performed more learning oriented actions. Corgnet (2012) studied how different acquaintance levels and money-sharing rules affected motivation for PA, finding that, to maintain intrinsic motivation, when working groups consist of friends it was better to have fixed sharing rules, while in teams with low acquaintanceship, making the amount of money to be received dependent on peer rating was a better sharing method.

A student's sense of fairness or (dis)comfort in PA is another emotional response in which contradictory results have been found. Some studies indicate that students perceive PA as unfair (Carvalho, 2012; McConlogue, 2010; Vanderhoven et al., 2015; Wen & Tsai, 2006; Willey & Gardner, 2010; Wilson et al., 2015) and

something they are not comfortable with (Carvalho, 2012; Vanderhoven et al., 2015). A sense of fairness in PA can lead to a positive acceptance of PA, greater comfort in PA, and less friendship bias marking (Carvalho, 2012; Panadero et al., 2013). Friction among teammates within a collaborative learning activity generated negative emotions (Carvalho, 2013). However, other studies indicate confidence in the fairness of PA (Falchikov, 1995; Hunter & Ross, 1996; Panadero et al., 2013), with students expressing comfort with it (Panadero et al., 2013). Part of the explanation for these contradictory results may lie in low or biased response rates (Wilson et al., 2015), lack of detail about the measurement of fairness (Falchikov, 1995; Hunter & Russ, 1996), or shortness of intervention (Panadero et al., 2013). In sum, PA can be a problematic assessment procedure in terms of perceived fairness by students.

Just as fairness has contradictory evidence, so does student emotion within PA. Hou and Cheng (2012) found PA had unclear effects on student emotional patterns. Cheng et al. (2014) found that the students who received more comments via peer feedback on YouTube, responded more to the comments and expressed more positive emotions to their peers' neutral or positive comments. On the other hand, students who received fewer responses showed more negative emotions to critical and even encouraging messages by their peers. Vanderhoven et al. (2015) found that students in an anonymous PA condition reported more positive feelings towards PA. Hence, it is difficult to derive robust conclusions from so little research other than that PA affects students' emotions. Students' perceptions of PA fairness and (dis)comfort are problems in the classroom unless PA addresses and overcomes these valid concerns.

4.3. Inter-personal aspects of PA

Social and inter-personal factors are important in PA, especially when PA is not anonymous (Cheng & Tsai, 2012). The importance of psychological safety for the

effective functioning of PA is now well established (van Gennip el al. 2009; van Gennip, Gijbels et al., 2010); Cheng and Tsai (2012) found that students with higher psychological safety chose deeper learning approaches to their online PA. When PA participants feel safe, they have greater confidence that the peer assessor will make a valuable contribution. Two interventions that enhanced psychological safety were seen in van Gennip, Segers et al. (2010), where a control group was compared to a PA training group with enhanced safety measures, and in van Gennip, (2012b) that compared a control group against two different PA interventions with the PA conditions surpassing the control group on psychological safety. These studies show that training students to perform PA can increase their perceptions of psychological safety.

PA with one's friends is assumed to be psychologically safe but also tends to lead to a friendship scoring bias. Two studies found students reported inflating PA scores in order to enhance relationship with peers (Carvalho, 2012; Harris & Brown, 2013), while Panadero et al. (2013) showed that students over-scored peers regardless of friendship level. Falchikov (1995) found that one of the least liked features of PA was marking friends, a reason why Hunter and Ross (1996) used students from more advanced courses to score second year music students. Hence, anonymity in PA may reduce peer pressure and fear of disapproval (Cheng & Tsai, 2012, Panadero et al., 2013; Vanderhoven et al., 2015).

Nonetheless, despite friendship bias, the relationships of peer marks to instructor grades are positively correlated and, thus, potentially valid (Falchikov & Goldfinch, 2000; Panadero et al., 2013). However, the direction and strength of friendship bias is not consistent; different rules of sharing the money awarded to the team for the group work (Corgnet, 2012) and different types of PA –just asking students to conduct unguided PA vs. Rubric PA– (Panadero et al., 2013) impact degree of bias. Social

aspects are crucial to PA, impacting the validity and meaning of PA. Three of the studies showed that students feel social pressure while performing PA when scoring their peers (Carvalho, 2012; Harris & Brown, 2013) or in comparison to a non-anonymous PA (Vanderhoven et al., 2015). This scoring bias perception is actually transmitted to the peer score in the form of inflation, sometimes regardless of friendship level, and has been decreased using rubrics (Panadero et al., 2013). Therefore, there seems to be a tendency to over-score friends in non-anonymous PA.

4.4. Cognitive aspects of PA

For PA to be an effective learning strategy, students must have confidence in themselves and their peers as evaluators (van Gennip et al., 2009). Generally, students report difficulties and problems in trusting that their peers can give competent assessments and feedback about their work (Harris & Brown, 2013; Lin et al., 2002; van Gennip, Segers et al., 2010), although this is not always consistently found (Li & Steckelberg, 2004). Students also express concern that it is not appropriate to do PA because “it is not their job” (Willey & Gardner, 2010; Wilson et al., 2015). The higher the trust in the other as assessor, the deeper the learning approaches towards PA (Cheng & Tsai, 2012). Therefore trust in the other is something we want to promote.

Nonetheless, efforts to induce greater trust among peers in PA have not always had positive impact: van Gennip, Segers et al. (2010) and Sluijsmans et al. (2002) found better performance and more development of PA skills in their PA condition in comparison to the control groups but these interventions had no effect on trust in either the self as assessor or the other as assessor. On the other hand, McMahon (2009) showed a major effect on trust in themselves as assessors through a very intensive intervention. Similarly, Topping et al. (2000) also found increased trust in PA when participants in their intensive intervention were given training in PA; the training

focused on developing experience with PA, using anonymous reporting initially, and reporting only positive aspects of performance. Additionally, greater expertise in the PA assessor is associated with the assessee having greater trust in a peer's assessment (Wen & Tsai, 2012; Willey & Gardner, 2010). Greater transparency about feedback also helps raise trust (van Gennip, Gijbels et al., 2010). In sum, trust in PA does not happen automatically and intensive practice and interaction seems to increase it.

Since PA, at least with a formative purpose, requires peers to collaborate with each other to conduct and accept evaluations, it is important that they have a shared understanding of what effective PA looks like, beyond mastery of techniques. To visualize this, when students have contrasting beliefs about the point of PA (e.g., one student believes that PA is just about scoring a peer work without further interaction, while another believes differences in opinion lead to interesting discussion in which lies the potential of PA), this is known as value diversity, as opposed to value congruence. The studies explored here showed that the more value congruence in PA (or the less value diversity), the more trust there is and the more valuable the PA results (van Gennip, 2012b; van Gennip, Gijbels, et al., 2010; van Gennip, Segers et al., 2010); deeper learning approaches are also activated. (Cheng & Tsai, 2012). Additionally, it has been found that, at the same time, PA interventions can increase value congruency (van Gennip, 2012b; van Gennip, Segers et al, 2010). Therefore, practice in PA must be a requirement for students to gain trust in the whole process; recent research support this idea (Panadero & Brown, 2015). Furthermore, students who share the belief that greater learning occurs through the PA process (i.e., interdependence), rather than valuing independent or self-reliance, have a greater trust in each other (van Gennip, Gijbels, et al., 2010) and deeper learning approaches (Cheng & Tsai, 2012), both of which have been shown to be important to the quality of PA.

However, whether value congruency and interdependence can be generated in anonymous, randomly-assigned products (e.g., online PA) contexts where the students barely interact is unknown, but it could be the case that because the students do not interact, they have differentiated goals (e.g. here I act as an assessor, here I am an assessee) and without interaction, the alignment of goals can prove rare. In contrast, it seems that value congruency and interdependence play an important role when intensive types of PA are implemented (e.g. face to face feedback) because the interaction of assessor-assessee can produce such shared goals.

In sum, while it seems like value diversity/congruence has some connection to PA, they have not been explored much further than van Gennip's work. One of the reasons is that, while value diversity/congruence are applicable in collaborative learning situations where the goals are shared (or at least they should be), in classroom PA such shared goals are much less frequently the case. Peer assessment practices can happen in isolated contexts (e.g., online anonymous PA) where the goals of assessors and assessees are divergent: the assessor evaluates and produces a score and/or feedback, while the assessee interprets and reacts to that information. It is actually the case that van Gennip et al. (2009) borrowed value diversity/congruence from collaborative learning research and the connection to the assessment field they established is not yet robust. Therefore, that connection needs to be clarified in further attempts to use value diversity/congruence in PA, by first clarifying how assessor and assessee's goals can be aligned.

5. **Discussion**

This review has shown that there are different ways to conceptualize and implement PA and that there are human and social effects of differing PA implementations on students that need further exploration. Additionally, the effect of

PA in individual differences is a completely unexplored area (e.g., how PA might be affected by different beliefs, personality), as a "one size fits for all" approach to PA is unlikely to have a uniform impact. Fortunately, since 2009, the PA field has focused more on these aspects and less on scoring accuracy, but this is an incomplete journey. Research indicates that more intensive PA implementations produce better human and social outcomes, especially as students gain a deeper understanding of the complexities of PA (McMahon, 2009; Topping et al., 2000). While more superficial approaches to PA (e.g., peer marking), are not wrong, they are not enough because (a) they do not guarantee enough peer feedback and other interactions that lead to more learning, (b) they do not explain why the grades have been awarded, and (c) non-disclosure can increase perceptions of unfairness and discomfort. These issues do not support the conclusion that summative approaches to PA should stop per se because there are grading-oriented approaches to PA (e.g., using rubrics to classify and comment on peer work) that can promote students' learning when well implemented (Harris & Brown, 2013). In sum, the conclusion is that superficial approaches that emphasize the scoring part of PA can be detrimental for interpersonal relationship in the classroom and do not guarantee more learning because peer scores without further feedback do not lead automatically to more learning. Therefore, if summative purposes are going to be used, it is better if the implementation is profound as this type can enhance reflection.

However, teachers need to understand the limited scope if such approaches are adopted. We cannot expect students to learn as much from anonymous peer-marking where the assessee receives little feedback as compared to PA where the assessees and assessors interact and rich feedback is provided. The more interaction and peer feedback we want to provide, the longer PA will take, the more psycho-social variables will need to be monitored, and the more PA will have to become a central part of the curriculum.

Therefore, if a teacher wants to implement a shallow approach to PA because the main goal is to have students score each other, he or she will have to be aware of the negative social consequences that peer marking and lack of peer feedback could bring.

Along this line, if a major goal of PA is to enhance learning, then greater focus on content accuracy (i.e., substantively appropriate information) becomes more important than scoring accuracy. Formative types of PA which focus on learning (Corgnet, 2012; Li & Steckelberg, 2004; Reinholz, 2015; Sluijsmans, Brand-Gruwel, & van Merriënboer, 2002; Van Zundert et al., 2010) usually require more social interaction as assessors provide more specific feedback or even interact face to face or via a computer with the assessees (Kollar & Fischer, 2010; Strijbos et al., 2009; van Gennip et al., 2009). Thus, interpersonal variables may be of greater relevance when PA research is focused on content and learning goals rather than simply on the consistency of student scoring with that of teachers or tutors.

This review has highlighted a large number of areas where insufficient research has been conducted. Within the field of intra-individual factors, more research is needed comparing different PA implementations as there seem to be differential effects. For example, research is needed into: (a) the role of goal orientation in PA activation, (b) the effect of PA formats on motivation and emotions, (c) the impact of anonymity and exposure, in both face-to-face and online contexts, upon emotion and motivation, (d) the effects of cultural differences, and (e) how the effect of classroom motivational climate (Alonso-Tapia & Fernandez, 2008) might affect PA (especially psychological safety, trust, fairness and (dis)comfort).

Although the nationality of the research samples have been included in Table 1, it cannot be concluded that the results reflect cultural or social group characteristics; rather the diversity indicates that international interest in PA is robust. Current research

has not considered a wide breadth of possible emotional responses to PA, suggesting, for example, that research with the broader range of emotions in the Academic Emotions Questionnaire might be useful (Pekrun, Goetz, Titz, & Perry, 2002).

Likewise, research into the inter-personal dimensions needs to consider: (a) the impact of friendship on PA scoring and content accuracy, while not limiting the evidence to self-reported data; (b) the effect of possible interventions to reduce student tendencies to over-score; (c) the impact on friendship when PA requires a negative evaluation; (d) the effect of different PA types on more general social classroom climate, (e) the impact of different qualities of peer feedback (e.g., positive vs. negative; deep vs. superficial) upon assessees' emotions, motivations, and their ability to process such feedback, and (f) the emotions and motivations of assessors' while giving different evaluations (e.g., positive vs. negative; low vs. high quality of peer work).

Research has shown that trust between peers in PA is not just a question of personality or attitude, but also depends on the competence of peers doing the evaluations. Hence, research is needed into the effect of assessor and assesse expertise and its influence on the quality of PA and the human and social impact.

This review also highlights some important implications for policy and practice. The first and more important is that we should not underestimate the interaction of PA and human and interpersonal variables. PA has an effect at the motivational and emotional level and can trigger powerful feelings such as unfairness and discomfort (like all other forms of educational assessment). Therefore, we need to implement PA in ways that minimize such negative effects. The studies reviewed here show that, the "deeper" the PA (i.e., more feedback and assessor-assessee interaction), the better results in the intra and inter-personal variables studied here. But this comes with the "cost" of taking more classroom time as students need to be taught and scaffolded as to

how to perform PA, and given the space and time for interaction. In other words, PA in such cases becomes a central part of the curriculum. This is the way PA reaches its full potential for enhancing students' self-regulation and learning. If teachers still prefer to use a summative approach (e.g., peer marking) to PA, they need to be aware of the possible tensions this could create, alongside possible decreases in student interest in PA: students' evaluation is important and if the PA approach is shallow, this will impact students' liking of the process.

Can PA still be used for summative purposes? Even though less preferably than formative uses, if we want to use PA for grading students, we need to be aware of the following: (a) it will take considerable time for the teacher to train and scaffold students to do this effectively, so it may not actually 'save time', (b) assessors need to provide a rationale for their score, (c) the use of assessment criteria (e.g., rubrics) will enhance the reliability and decrease friendship bias, and (d) the use of anonymity has to be considered carefully in terms of the learning benefits, if any, that it could produce (i.e., anonymity may allow assessors to focus more on the content, than be concerned by who the assessee is and how he/she might take their assessment). In sum, PA implementations should be planned thoughtfully, not as a trial and error adventure.

These facts have at least three important implications. First, assessment for learning should be more critical of PA stances that merely award grades or scores without explanation; such approaches at least require students to be involved in assessment and, thus, have some value, but learning and self-regulated learning are not central to them. Therefore, assessment for learning proponents need to clearly position themselves against the indiscriminate use of PA because shallow implementation might produce more harm than good. Second, pre-service teachers and in-service teachers should be trained in PA implementation and, additionally, have experience themselves

within PA as both assessors and assessees as recent research has shown this is crucial for teachers' future use of PA in their classrooms (Panadero & Brown, 2015). Special emphasis should be given to the human and social PA factors analysed here which are of major interest for the classroom climate. And, third, in terms of the psychometric analysis of PA, researchers need to move the focus from grade/scoring accuracy towards content accuracy, which is more beneficial for PA promotion of learning and self-regulated learning.

6. Conclusion

This chapter began with a well-known quote: "With great power should also come great responsibility". Peer assessment is an educational practice with strong social and human effects. For the assessor, they must be aware that their feedback and grade will affect the assessee. But researchers should also take note: for a long time the field has claimed that PA is a reliable source of information in terms of scoring accuracy. Now, we need to start exploring what happens in PA interactions and how it affects learning, not only the cognitive side of it (how much students learn), but also the motivational, emotional, and interpersonal aspects. Otherwise we will just continue trapped in our own research spider web, condemned to research the same topic (scoring accuracy) over and over again.

References

Alonso-Tapia, J., & Fernandez, B. (2008). Development and initial validation of the classroom motivational climate questionnaire (CMCQ). *Psicothema, 20*(4), 883-889.

Bryant, D. A., & Carless, D. R. (2010). Peer assessment in a test-dominated setting: Empowering, boring or facilitating examination preparation? *Educational Research for Policy and Practice, 9*(1), 3-15. doi: 10.1007/s10671-009-9077-2

Carvalho, A. (2012). Students' perceptions of fairness in peer assessment: Evidence from a problem-based learning course. *Teaching in Higher Education, 18*(5), 491-505. doi: 10.1080/13562517.2012.753051

Cheng, K. H., Hou, H. T., & Wu, S. Y. (2014). Exploring students' emotional responses and participation in an online peer assessment activity: A case study. *Interactive Learning Environments, 22*(3), 271-287. doi: 10.1080/10494820.2011.649766

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*(4), 599-618.

Corgnet, B. (2012). Peer evaluations and team performance: When friends do worse than strangers. *Economic Inquiry, 50*(1), 171-181. doi: 10.1111/j.1465-7295.2010.00354.x

Dennis, I., & Newstead, S. (1994). The strange case of the disappearing sex bias. *Assessment and Evaluation in Higher Education, 19*, 49-56.

Dixon, H. R., Hawe, E., & Parr, J. (2011). Enacting Assessment for Learning: The beliefs practice nexus. *Assessment in Education: Principles, Policy & Practice, 18*(4), 365-379. doi: 10.1080/0969594X.2010.526587

Dochy, F. (2006). A guide for writing scholarly articles or reviews for the Educational Research Review. *Educational Research Review*. http://www.journals.elsevier.com/educational-research-review/

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*(3), 331-350. doi: 10.1080/03075079912331379935

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education & Training International, 32*(2), 175-187. doi: 10.1080/1355800950320212

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.

Falchikov, N., & Magin, D. (1997). Detecting gender bias in peer marking of students' group process work. *Assessment & Evaluation In Higher Education, 22*(4), 385-396. doi: 10.1080/0260293970220403

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*(0), 101-111. doi: 10.1016/j.tate.2013.07.008

Hou, H. T., & Cheng, K. H. (2012). Analyzing the latent emotional transfer pattern (LETP) of a learning community in an online peer-assessment activity. *British Journal of Educational Technology, 43*(4), E113-E116. doi: 10.1111/j.1467-8535.2012.01301.x

Hunter, D., & Russ, M. (1996). Peer assessment in performance studies. *British Journal of Music Education, 13*(01), 67-78. doi: doi:10.1017/S0265051700002953

Hwang, G. J., Hung, C. M., & Chen, N. S. (2014). Improving learning achievements, motivations and problem-solving skills through a peer assessment-based game development approach. *Educational Technology Research and Development, 62*(2), 129-145. doi: 10.1007/s11423-013-9320-7

Karaca, E. (2009). An evaluation of teacher trainees' opinions of the peer assessment in terms of some variables. *World Applied Sciences Journal, 6*(1), 123-128.

Kim, M., & Ryu, J. (2013). The development and implementation of a web-based formative peer assessment system for enhancing students' metacognitive awareness and performance in ill-structured tasks. *Educational Technology Research and Development, 61*(4), 549-561. doi: 10.1007/s11423-012-9266-1

Koc, C. (2011). The views of prospective class teachers about peer assessment in teaching practice. *Educational Sciences: Theory & Practice, 11*(4), 1979-1989.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*(4), 344-348. doi: 10.1016/j.learninstruc.2009.08.005

Lee, S., & Ditko, S. (1962, August 15). Spider Man. *Amazing Fantasy, 15*, 1-11.

Li, L., & Steckelberg, A. (2004). *Using peer feedback to enhance student meaningful learning*. Paper presented at the Conference of the Association for Educational Communications and Technology, Chicago, USA.

Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2002). Student attitudes toward networked peer assessment: Case studies of undergraduate students and senior high school students. *International Journal of Instructional Media, 29*(2), 241-254.

Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education, 7*(1), 51-62. doi: 10.1177/1469787406061148

Lui, A., & Andrade, H. (2014). Student peer assessment. In R. Gunstone (Ed.), *Encyclopedia of Science Education* (pp. 1-3): Springer Netherlands.

Magin, D. (2001). Reciprocity as a source of bias in multiple peer assessment of group work. *Studies in Higher Education, 26*(1), 53-63. doi: 10.1080/03075070020030715

McConlogue, T. (2010). But is it fair? Developing students' understanding of grading complex written work through peer assessment. *Assessment & Evaluation In Higher Education, 37*(1), 113-123. doi: 10.1080/02602938.2010.515010

McMahon, T. (2009). Combining peer-assessment with negotiated learning activities on a day-release undergraduate-level certificate course (ECTS level 3). *Assessment & Evaluation In Higher Education, 35*(2), 223-239. doi: 10.1080/02602930902795919

Montero Garcia-Celay, I., & León, O. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology, 7*(3), 847-862.

Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation In Higher Education, 21*(3), 239-250. doi: 10.1080/0260293960210304

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology, 11*(2), 551-576. doi: 10.14204/ejrep.30.12200

Panadero, E., & Brown, G. T. L. (2015). *Teachers' reasons for using peer assessment: A survey self-report of Spanish teachers.* Manuscript submitted for publication.

Panadero, E., & Brown, G. T. L. (2015). *Higher education teachers' assessment practices: Formative espoused but not yet fully implemented*. Paper presented at the Fifth Assessment in Higher Education Conference 2015, Birmingham, (UK).

Panadero, E., Brown, G. T. L., & Courtney, M. G. R. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in*

Education: Principles, Policy & Practice, 21(3), 365-383. doi: 10.1080/0969594X.2014.919247

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., & Strijbos, J. W. (2013). The impact of a rubric and friendship on construct validity of peer assessment, perceived fairness and comfort, and performance. *Studies In Educational Evaluation, 39*(4), 195-203. doi: 10.1016/j.stueduc.2013.10.005

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91-105. doi: 10.1207/S15326985EP3702_4

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

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1 - 31.

Sluijsmans, D. M. A., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment & Evaluation In Higher Education, 27*(5), 443-454. doi: 10.1080/0260293022000009311

Stanier, L. (1997). Peer assessment and group work as vehicles for student empowerment: A module evaluation. *Journal of Geography in Higher Education, 21*(1), 95-98. doi: 10.1080/03098269708725413

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 C. Mourlas, N. Tsianos & P. Germanakos (Eds.), *Cognitive and Emotional Processes in Web-Based*

Education: Integrating Human Factors and Personalization (pp. 375-395). Hersey, PA: IGI Global.

Taras, M. (2005). Assessment -summative and formative- some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466-478.

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276.

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (Vol. 1, pp. 55-87): Springer Netherlands.

Topping, K. J. (2012). Peers as a source of formative and summative assessment. In J. H. McMillan (Ed.), *Handbook of research on classroom assessment* (pp. 395-412). Los Angeles: SAGE.

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation In Higher Education, 25*(2), 149-169. doi: 10.1080/713611428

van den Bossche, P., Gijselaers, W. H., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments. *Small Group Research, 37*(5), 490-521. doi: 10.1177/1046496406292938

van Gennip, N. (2012a). Arranging peer assessment: The role of interpersonal beliefs *Assessing together. Peer assessment from an interpersonal perspective* (Vol. PhD, pp. 45-57): Universiteit Leiden.

van Gennip, N. (2012b). *Assessing together. Peer assessment from an interpersonal perspective.* (PhD), Universiteit Leiden.

van Gennip, N., Gijbels, D., Segers, M., & Tillema, H. H. (2010). Reactions to 360º feedback: The role of trust and trust-related variables. *International Journal of Human Resources Development and Management, 10*(4), 362-379. doi: 10.1504/IJHRDM.2010.036088

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*(1), 41-54. 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., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270-279. doi: 10.1016/j.learninstruc.2009.08.004

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: http://dx.doi.org/10.1016/j.compedu.2014.10.001

Wen, M. L., & Tsai, C. C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 51*(1), 27-44. doi: 10.1007/s10734-004-6375-8

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537-548. doi: 10.1080/014119296022050

Willey, K., & Gardner, A. (2010). Investigating the capacity of self and peer assessment activities to engage students and promote learning. *European Journal of Engineering Education, 35*(4), 429-443. doi: 10.1080/03043797.2010.490577