



Universiteit
Leiden
The Netherlands

Assessing together : Peer assessment from an interpersonal perspective

Gennip, A.E. van

Citation

Gennip, A. E. van. (2012, October 23). *Assessing together : Peer assessment from an interpersonal perspective*. Retrieved from <https://hdl.handle.net/1887/20012>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/20012>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/20012> holds various files of this Leiden University dissertation.

Author: Gennip, Anna Eva (Nanine) van

Title: Assessing together. Peer assessment from an interpersonal perspective

Issue Date: 2012-10-23

CHAPTER 4

Peer Assessment as a Collaborative Learning Activity: The Role of Interpersonal Beliefs and Conceptions⁷

The present study examined the role of interpersonal beliefs (psychological safety, value congruency, interdependence, and trust) and conceptions of peer assessment in vocational education. An intervention was conducted (N = 45) with a control group (N = 17), which indicated change in psychological safety, value congruency, and trust in the peer as an assessor. Furthermore, when comparing the intervention and control group, peer assessment contributed to psychological safety and higher value congruency. Perceived learning was predicted by value congruency and conceptions. Conceptions were predicted by psychological safety, value congruency, and trust in the self and in the peer as an assessor.

1 Introduction

Many studies indicate that student learning is positively influenced by assessment (Black & Wiliam, 1998; Kennedy, Chang, Fok, & Yu, 2008; Pellegrino, Chudowsky, & Glaser, 2001). Assessment informs students about their strengths and weaknesses and indicates the next steps to take in the learning process. One important condition for assessment to support student learning is the active involvement in the assessment process on the part of students themselves (Black & Wiliam, 1998). As a result, students can make an active contribution to their own knowledge construction, which is beneficial to learning outcomes (Sluijsmans, 2002). This view has become known as the ‘assessment for learning’ position (Black & Wiliam, 1998).

A frequently adopted assessment method in which students are actively involved in the appreciation and appraisal of learning is peer assessment, as this is

⁷ Based on 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.

closely embedded in and aligned with students' efforts during the instructional process (Shepard, 2000). In peer assessment students learn from each other by means of receiving and giving feedback. Topping (1998) defines peer assessment as «Peer assessment is an arrangement in which individuals consider the amount, level, value, worth, quality or success of the products or outcomes of learning of peers of similar status» (p. 250). However, several studies have shown that the effects of peer assessment are diverse: for example, peer assessment is said to be beneficial to the learning process (Davies, 2002). More specifically, it has been found that peer assessment (together with self- and co-assessment) does help students to develop certain skills in the areas of, for example, communication, self-evaluation, observation, and self-criticism (Dochy & McDowell, 1997).

1.1 Peer assessment is a social process

Confirming the diverse picture regarding the effects of peer assessment the literature reviews by Dochy, Segers, and Sluijsmans (1999) and Topping (2003) showed that although studies on peer assessment seem to have found positive effects in general, the results remain inconclusive. More recently, Van Gennip, Segers, and Tillema (2009) conducted a literature review on empirical studies in higher and professional education that measured learning gains in peer assessment settings. For the period 1990-2007 they were able to identify only fifteen studies. This result indicates that there is still very little evidence on the effects of peer assessment on student learning. Moreover, regarding the effects found, the diverse picture that emerged from earlier review studies was confirmed by Van Gennip et al. (2009). One of the reasons for the inconclusive results might be that in some studies the variety in assessment interventions may have been more beneficial to learning in some settings than in others. For example, differences arose in face-to-face versus distance assessment, and confidential versus public peer assessment formats. Because of a lack of research relating features of the peer assessment setting to learning gains, it is hardly possible to draw conclusions at this point (Van Zundert, Sluijsmans, & Van Merriënboer, 2010).

In this respect, it is surprising that hardly any study has addressed the interpersonal context in which the peer assessment intervention took place. Reviewing the nature of peer assessment we find that it is an inherently social process in which students, by assessing each other, learn with and from each other as peers. It is especially in the collaborative definition and/or discussion of the criteria and standards for achievements to be appraised (see Van Steendam, Rijlaarsdam, Ser-cu, & Van den Bergh, 2010), and the nature of the feedback (see Cho & MacArthur, 2010; Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; Strijbos, Narciss, & Dünnebier, 2010), that learning takes place. As a consequence the question what constitutes beneficial peer assessment is raised and, in particular, how interper-

sonal beliefs are interrelated, since one might contend that it is in the social nature of the appraisal process that students come to accept each other's assessments and learn from it.

1.2 Peer assessment as a learning intervention

It is clear that it takes more than bringing students together to make learning a collaborative activity. During the past decades research on team learning has highlighted the importance of the interpersonal context in which team learning takes place. There is evidence that the development of and the interplay between interpersonal beliefs affect the outcomes of a collaborative learning activity (Edmondson, 1999; Van den Bossche, Gijssels, Segers, & Kirschner, 2006). Several interpersonal beliefs have been identified as important for team work. Unanimity of opinion about the team's task and mission (high value congruency), a belief that the environment is safe for interpersonal risks, that is, group members feel safe enough to say, do, and ask what they think is good (psychological safety), and a feeling of mutual dependence according to the task (interdependence) all proved important for learning, information sharing, good communication, and a good team performance (Edmondson, 1999; Jehn, Northcraft, & Neale, 1999; Van der Vegt, Emans, & Van de Vliert, 1998).

Within the peer assessment literature several authors have referred to the relevance of interpersonal beliefs as well. Topping (2003), for example, theorises: "Peer assessments might be partly determined by: friendship bonds, enmity or other power processes, group popularity levels of individuals, perception of criticism as socially uncomfortable or even socially rejecting and inviting reciprocation, or collusion leading to lack of differentiation" (p. 67). The studies by Dochy et al. (1999), Falchikov (1995), and Sluijsmans, Brand-Gruwel, and Van Merriënboer (2002) refer to various problems that might arise given the social context of peer assessment. They mention students' hostility towards peer assessment when they first experience it, a lack of trust in the self and the other as assessors, and friendship marking, where peers give their friends higher marks than others regardless of performance. Despite the various indications that interpersonal beliefs might play a significant role within peer assessment, these have to date hardly been studied in a systematic way (as shown by Van Gennip et al., 2009). One study (Stanier, 1997) was found that referred to the relevance of interpersonal beliefs, that is, how students conceive peer assessment as a learning experience. In Stanier's (1997) study, students reported that they enjoyed working in groups, there were not many personality clashes, they were working together on a task, and that they thought their performance improved by working with others. These findings refer to how students perceive the interpersonal context, or more precisely, psychological safety and interdependence. Additionally, 40% indicated that peer assessment was an

uncomfortable experience. However, with respect to perceived learning gains the majority of students (74%) stated that peer assessment was an awareness-raising experience which stimulated them to think about the quality of their peers' work (98%).

Although the studies reviewed offer some interesting findings with regard to peer assessment effects, they hardly provide empirical evidence on the nature of the peer assessment setting, that is, its interpersonal aspects of the setting contributing to learning. Therefore, it might be relevant to gauge the change in student perceptions with regard to both interpersonal beliefs and students' conceptions of peer assessment as a tool to measure learning, as this results from experiencing this mode of assessment. Moreover, we need to establish more clearly how students' perceptions of the interpersonal beliefs, their conceptions of peer assessment, and learning gains relate to each other. This study will focus on the question how peer assessment as an intervention influences students' perceptions of the interpersonal beliefs, and their conceptions of peer assessment. In addition, it addresses the relation between interpersonal beliefs and conceptions of peer assessment with regard to learning gains.

1.3 Interpersonal beliefs and peer assessment

Several interpersonal beliefs come into play when arranging a collaborative or peer-based intervention, such as psychological safety, trust, value congruency, and interdependence.

1.3.1 Psychological safety

Psychological safety can be described as a shared belief that it is safe to take interpersonal risks in a group of people. As Edmondson (1999, p. 354) says «The term is meant to suggest neither a careless sense of permissiveness, nor an unrelentingly positive affect but rather a sense of confidence that the team will not embarrass, reject, or punish someone for speaking up». Until now psychological safety has not been an explicit issue in peer assessment studies. Implicitly, however, it has been acknowledged that peers have a tendency to assess on the basis of issues such as friendship and uniformity (Dochy et al., 1999). Thus, the social context is recognised as an influential factor in peer assessment.

The idea that psychological safety may influence the learning effects of peer assessment has arisen because of the positive impact of psychological safety on learning and group effectiveness that was found in several studies (Edmondson, 1999; Van den Bossche et al., 2006). Psychological safety, for example, prevents teams from perceiving differences in viewpoints as disagreements, and creates room for differences to be seen as opportunities to frame a problem. As a result,

psychological safety improves performance – not directly, but through facilitating the appropriate behavior leading to better performance (Edmondson, 1999).

Because peer assessment is fundamentally a social process, with feedback given to and received from others as the core activity, we hypothesised that positive appraisals on psychological safety will enhance the process of peer assessment. In their review Dochy et al. (1999) found that perceived openness, as an indicator of psychological safety, was fundamental to a fair assessment. Therefore, it might be expected that when peers perceive their environment as safe for interpersonal risk-taking they will be less prone to, for example, friendship marking, and will put effort into achieving a fair peer assessment process. Psychological safety, we contend, is a precondition for appraisal in a task-oriented and goal-directed way – a prime condition for peer assessment to support student learning (Assessment Reform Group, 2006).

1.3.2 Trust in the self and the peer as assessor

Several studies note that students feel uncomfortable criticising each other's work, or find it difficult to rate their peers (Topping, Smith, Swanson, & Elliot, 2000). This is partly a result of the novelty of peer assessment in education. Staff, but students as well, generally have little experience with this form of assessment. Ballantyne, Hughes, and Mylonas (2002) refer to various studies indicating that students feel assessment to be the responsibility of teachers, who are recognised as the experts on appraising learning. They conclude that students lack confidence in both their and their peers' abilities as assessors. For example, Orsmond and Merry's (1996) results suggest that many of the students were sceptical about the added value of peer comments. McDowell (1995) found that students expressed concerns about their ability to provide constructive feedback and mark fairly.

The influence of confidence or trust in both the self and the other in relation to appraising learning effects has until now hardly been addressed in empirical studies on (peer) assessment (Tillema, 2009). Therefore, Topping (1998) as well as Falchikov and Goldfinch (2000) suggest that future research should focus on the (perceived) quality of peers as assessors. In other words, the trust that students have in their and their peers' ability as assessors could influence perceived learning from peer assessment.

1.3.3 Value congruency

Value congruency is defined as the similarity in opinion about what a team's task, goal or mission should be (Jehn et al., 1999). In other words, value congruency is not about individual interest per se, but about whether group members agree on what is important for the group in order to perform well. Jehn et al. (1999) have shown that value congruency in teams should be high in order to be effective. In addition, Van Gennip, Van den Bossche, Gijsselaers, and Segers (2004) showed that

work teams performed better when value congruency was high. Integrating different perspectives and developing a shared understanding is crucial for teams to perform well (Van den Bossche et al., 2006). The importance of developing a shared understanding has been widely argued in reviews on peer assessment (Dochy et al., 1999; Falchikov & Goldfinch, 2000; Topping, 1998; Topping, 2003).

The necessity of a shared understanding is especially stressed with respect to assessment purposes, objectives, criteria, and standards (Jehn et al., 1999). Using their knowledge and skills to review, clarify, and evaluate other people's work is a cognitively demanding task for students involved in peer assessment. They are required not only to consider the objectives and purposes of the assessment task (Boud, 1995; Topping et al., 2000), but also to contemplate the questions of which criteria to use for assessing the work, and which standards to employ in order to identify a good or poor piece of work (Searby & Ewers, 1997). Because of the importance of generating assessment criteria and standards to enhance the learning effect of peer assessment, Boud (1995) and Ballantyne et al. (2002) recommend procedures to ensure that all elements important for an appraisal of (learning) outcomes are included in the assessment criteria. Therefore, criteria should be amended and shared where necessary in order to reach optimum understanding between peers. Given all this, we contend that low value congruency will have a positive influence on peer assessment for learning.

1.3.4 *Interdependence*

Interdependence between group members has been widely studied as an interpersonal belief in education (Johnson & Johnson, 1989; Mesch, Marvin, Johnson, & Johnson, 1988) as well as in organisations (Wageman, 1995). A distinction can be made between outcome interdependence and task interdependence (Van der Vegt et al., 1998). *Outcome interdependence* is defined as the extent to which team members believe that their personal benefits and costs depend on successful goal attainment by other team members (Van der Vegt et al., 1998). *Task interdependence* (initiated and received) refers to the interconnections between tasks that cause the performance of one specific piece of work to depend on the completion of certain other pieces of work (Van der Vegt et al., 1998). Studies have shown that task interdependence leads to more communication, helping, and information sharing than individualistic tasks (Crawford & Gordon, 1972; Johnson, 1973).

When peer assessment is implemented as an intervention to support learning it is meant to be an integrated part of a collaborative learning process. Task interdependence can then function as the «glue that holds the members together» (Sluijsmans, 2002, p. 2), that is, connects group members for the purpose of the task. Peer assessment implies that multiple perspectives on a task are made explicit, and requires students to be individually responsible for an active contribution to the group task. We therefore contend that learning from peer assessment occurs

when peers perceive interdependence (i.e., they see themselves as linked to each other on a task) to the extent that the assessment cannot be performed successfully unless everyone participates in a responsible way. In this study we will focus on this specific aspect of group interdependence.

1.4 Students' conceptions of peer assessment

During the past decades a number of studies have been conducted on students' conceptions of assessment, indicating their importance for the acceptance and validity of assessments. Thompson (1992, p. 130) considers conceptions as "a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like." Furthermore, student conceptions represent different categories of ideas that are at the bottom of students' descriptions of how educational matters are experienced (Pratt, 1992). This implies that conceptions can be described as a framework through which a student views, interprets, and interacts with the learning environment (Marton, 1981). There is a growing body of research indicating that conceptions of assessment are of significant importance for student learning (Hirschfeld & Brown, 2009). It is argued that people generally perform better the more positive their conceptions regarding a task are (Brown, Irving, Peterson, & Hirschfeld, 2009), partly because their positive conceptions make them feel more competent at the task.

There are generally two approaches in the research on conceptions of assessment. On the one hand, conceptions of the *purpose* of assessment have been studied (Brown et al., 2009). From this perspective, Brown et al. (2009) made a distinction between four different conceptions: assessment improves learning, assessment makes students accountable, assessment is negative and irrelevant, and assessment is liked. On the other hand, research can be found on conceptions of the *characteristics* of assessment. A relevant study was performed by Crossman (2004) revealing that conceptions of assessment are only partly determined by prior experiences; anxiety, student notions of relevance, and student-teacher relationships were among those aspects also found to be influential (Crossman, 2004). Especially the last factor, student-teacher relationships, is interesting within the context of the current study as an interpersonal belief that can influence the conceptions of assessment. In a peer assessment setting, which is by definition social and interpersonal, the perceived relationships among peers could influence existing conceptions and therefore affect the outcomes of the peer assessment.

In the context of peer assessment there are a few studies that refer to students' conceptions. The review study by Dochy et al. (1999) describes earlier research addressing the question of how students perceive fairness of peer assessment. This review shows that students perceive peer assessment as a sufficiently fair process, and described openness and clarity as fundamental to a fair assess-

ment. Topping (2003) indicates that assessors and assessees might experience initially anxiety about the peer assessment process. Further, Sluijsmans et al. (2002) mention the hostility that students show towards peer assessment when they first experience it. Insufficient introduction to the process of peer assessment may be an important reason. In this respect, Dochy et al. (1999) refer to earlier studies revealing that students' conceptions of peer assessment generally change for the better as they gain more experience with this mode of assessment.

1.5 Perceived learning as dependent variable

Considering the 'perceived' character of the independent variables, we chose to use 'perceived' learning as a dependent measure, in order to keep the research design coherent. In social psychology, self-recording of one's own learning is a common measured variable (see Bandura, 1986). In previous research, student perceptions of learning in a course correlated much higher with student ratings of instruction than did differences in pretest and posttest scores (O'Connell & Dickinson, 1993), and students' perceived learning correlated highly with perceived teaching effectiveness (Ryan & Harrison, 1995).

Additionally, studies in the area of assessment (Atwater & Brett, 2005) point to the importance of people's perceived improvement in functioning as a result of assessment (in this case 360 degree feedback⁸), because it has great influence on how managers will ultimately respond. As Atwater and Brett (2005) argue "The immediate reactions managers have to 360 degree feedback are important because the ways an individual 'feels' about and reacts to the feedback may influence how or whether the individual changes his or her behavior in response to the feedback" (p. 533).

1.6 Research questions – Hypotheses

Although peer assessment is a collaborative process in which interpersonal beliefs play a role, to date hardly any attention has been paid to the role of these beliefs in studies on peer assessment interventions. In the present study peer assessment intervention was interpreted as an interactional process. Therefore, the first aim was to measure how the intervention contributed to a change in interpersonal beliefs (psychological safety, interdependence, value congruency, and trust), that is, how these beliefs were affected by peer assessment as a process. Hence, the first research question was: "Does participating in a peer assessment intervention result in a change in perceptions of interpersonal beliefs (psychological safety, interdependence, value congruency and trust) and conceptions of peer assessment

⁸ 360° feedback is an individual assessment, often used in performance-oriented environments. It involves multiple raters, often including the participants themselves.

over time?” Pre-perceptions (prior to the peer assessment intervention) with post-perceptions (measured after the intervention) were compared, as well as a peer assessment condition with a control group. It was expected peer assessment intervention to lead to higher scores on psychological safety, trust, and interdependence, and to higher scores on value congruency (Hypothesis 1).

A second aim of the present study was to explore how the interpersonal beliefs are related to students’ conceptions of peer assessment and to students’ learning gains (as perceived by the students, and expressed by their performance). The respective research question was: “What is the impact of interpersonal beliefs and conceptions of peer assessment in relation with (perceived) learning in a peer assessment setting?” In other words: “Is there a relation between (perceived) learning, interpersonal beliefs and conceptions?” It was expected that interpersonal beliefs play a significant role in peer assessment, and influence conceptions of peer assessment as well as perceived learning (Hypothesis 2).

It was further hypothesised that conceptions act as a mediating variable between interpersonal beliefs and perceived learning (Hypothesis 3). The conceptual model is shown in Figure 1.

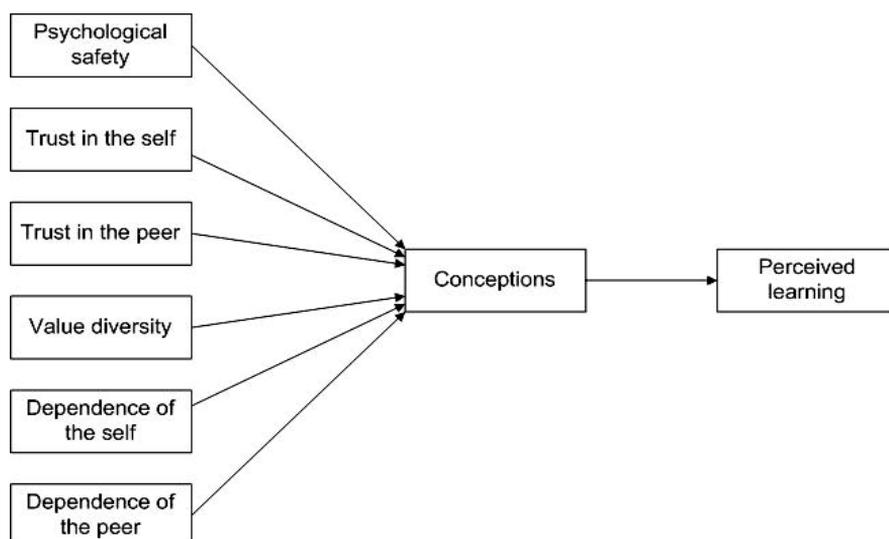


Figure 1

The conceptual model of the present study

2 Method

2.1 Participants

Participants in the study were 62 third-year male students in Dutch secondary vocational education. Their age ranged from 16 to 19 years. Students worked within a project-based course, offered at one large institute of technical vocational education consisting entirely of male students and focused on teaching detailed technical skills in metal work and electronics. The population of students taking the course was divided into 17 groups of three to five students, who worked together on a project for six weeks. Their project was to design and construct a robot artefact: a moving device with pneumatic and hydraulic elements.

2.2 Design

The study consisted of two steps. First (Research Question 1), the peer assessment intervention was studied through a factorial within-subject-change experimental design (Winer, 1984). This set-up was used to establish differences in perceptions within the same student (i.e., it served as its own control) and was labelled the 'experimental group' (N = 45). Interpersonal beliefs were measured at the beginning and the end of the course. Second (Research Question 2), a baseline condition was added so that we could compare this group with students who were not involved in the peer assessment intervention but took the same course and were assessed by their own teacher (N = 17). Students were randomly assigned to the control and experimental groups.

The experimental group used peer assessment to appraise the quality of the project product (i.e., a robot artefact). Teacher marks were collected for all project products, so we were able to compare control group teacher marks and experimental group teacher marks. These marks, however, were not given out to the students.

2.3 Procedure

At the beginning of the school year students were randomly divided over six classes. The experimental group consisted of twelve project teams (four classes, with three to four students per project team), each collaboratively working on their robot artefact. The control group consisted of five project teams (two classes, with three to four students per project team).

2.3.1 *Experimental group*

At the start of the project the experimental group received a two-hour instruction on peer assessment: the concept of peer assessment was explained, interaction

strategies between peers were discussed with the students, and eleven appraisal criteria were formulated to be used in the appraisal of each other's work. During the project separate groups worked on their artefacts and also received instruction in plenary sessions. At the end of the six-week project all groups gave a presentation on their task. After this presentation the robot artefact of each project team within the experimental condition was assessed by the peers not belonging to that project group, as well as by their teachers (for the research purposes mentioned). This was done on a form listing all eleven criteria. These criteria had previously been formulated by the students themselves (with some coaching from the teachers and the researcher) during the instruction at the beginning of the project (See Appendix 1 for the criteria form, translated from the Dutch language). The criteria were rated with 1 (good) or 0 (poor). Completed criteria forms were collected by the researcher, who calculated the ratings and returned these ratings to the project groups a week later.

All project groups were assessed by individual peers. In other words, all students assessed on an individual basis the other project groups in their class as a group. A questionnaire including all scales for interpersonal beliefs to be measured was distributed to the experimental groups both at the start of the project (pre-test), and after the presentation (posttest).

2.3.2 *Control group*

In the control group there was no training, and the project teams were assessed only by the teacher, on the same criteria as the students in the experimental group. The control group was used to benchmark the scores of the students in the experimental groups at the end of the project. In the control group condition, the questionnaire was only distributed at the end of the project (because there was no peer assessment intervention). Only those scales not directly related to the peer assessment intervention were included in the control group questionnaire (i.e., value congruency, psychological safety, perceived learning, trust in the self as assessor, interdependence). The intervention-related scales (i.e., trust in the peer as assessor, conceptions of peer assessment) were not relevant for the control group and therefore not administered. Students in the control group completed their questionnaire after teacher assessment, but before teacher marks had been given out.

2.4 Instruments

Variables in this study were the student perceptions; these were measured by means of a questionnaire, with most scales taken from existing, validated questionnaires. The questionnaire administered had been piloted in a secondary-vocational education setting. All items were measured using 5-point Likert scales, and anchored by 1 (totally true) and 5 (totally untrue). Reliabilities of all scales are

shown in Table 1, for the entire sample as well as for the experimental and control groups separately.

Table 1
Scales/subscales of the questionnaire and Cronbach's α reliability coefficients

		Total sample ($N = 62$)	Experimental group ($N = 45$)	Control group ($N = 17$)
	Number of items	α	α	α
Psychological safety	7	.57	.56	.50
Trust				
Trust in the self as assessor	4	.64	.65	.60
Trust in the peer as assessor	4	.71	.71	--
Value congruency	6	.85	.82	.83
Interdependence				
Dependence of the self	4	.83	.88	.61
Dependence of the peer	4	.78	.72	.88
Conceptions	10	.87	.87	--
Perceived learning	3	.78	.79	.79

2.4.1 Psychological safety

This scale measures the degree in which students perceive their group as safe for interpersonal risk taking. The scale has been derived from Edmondson (1999) and consists of seven items. A sample item is: "It is easy to ask my peers for help." However, reliability is rather low (Cronbach's $\alpha = .57$).

2.4.2 Trust

This scale measures trust in the self and the peer as assessor. We used an adapted version of the Assessment Skill scale by Sluijsmans et al. (2002), which we also expanded to measure trust in the peer as an assessor (e.g., "My peers are good at giving feedback"). Both subscales include four items. Reliabilities for Trust in the Self as Assessor (Cronbach's $\alpha = .64$) as well as for Trust in the Peer as Assessor (Cronbach's $\alpha = .71$) are acceptable.

2.4.3 Value congruency

The scale measuring value congruency addresses the perceived similarities between group members on group task, and goal or mission. It was adopted from a study by Jehn et al. (1999), and consists of six items. Sample items are: "The group as a whole has one single goal" and "Group members agree on what is important for the group" (Cronbach's $\alpha = .85$).

2.4.4 *Interdependence*

This scale measures two aspects of interdependence, namely dependence of the self, and dependence of the peer(s). Both subscales were based on scales developed by Van der Vegt et al. (1998) and consist of four items each. Sample items are: "I depend on my peers for information and advice" and "My peers depend on me for information and advice", respectively. Reliability values for Dependence of the Self (Cronbach's $\alpha = .83$) and Dependence of the Peer (Cronbach's $\alpha = .78$) are acceptable.

2.4.5 *Conceptions of peer assessment*

The scale measuring conceptions of peer assessment consists of 10 items. It is a shortened version of a questionnaire developed by Sluijsmans et al. (2002). Sample items are: "Peer assessment is useful" and "You have to learn how to assess your peers" (Cronbach's $\alpha = .87$).

2.4.6 *Perceived learning*

We newly developed the three-item scale measuring perceived learning (Cronbach's $\alpha = .78$). It measures perceived learning gains resulting from partaking in the intervention when it comes to having one's own product appraised by peers, and appraising peers' products oneself. The items are: "Assessing my peers made it easier to make my own product", "Assessing each other was a good practice for me to make my own product", and "Assessing each other taught me to look critically at my own product."

2.5 **Data analysis**

In the analysis of the data descriptive and correlation analyses were conducted first. Then, in order to answer the first research question, three analyses were performed. First, a paired-sample *t*-test for difference of means of the pretest and posttest within the experimental group was performed, in order to detect changes in student beliefs. Second, as a benchmark the posttest data from the experimental group were compared with those from the control group. Finally, a logistic regression was carried out to test whether there was a relation between group (experimental versus control) and the independent variables. To further analyse these findings, a one-way analysis of variance was carried out to test for differences between specific interpersonal beliefs in the experimental and control groups.

To answer the second research question we used hierarchical regression analysis in order to test the supposed mediating role of conceptions of peer assessment between interpersonal beliefs and perceived learning. This regression model consisted of three steps. In a first step we tested the effect of all independent variables on conceptions. The second and third steps contained all the separate independent

variables, such as predictors of perceived learning, alternatively excluding and including conceptions as a predictor variable.

3 Results

Means, standard deviations, and Pearson correlations of variables measured in the experimental group are presented in Table 2. Regarding the relation between perceived learning and performance ratings (peer and teacher marks in both groups) the results of the correlation analysis showed that perceived learning (by students) was not related to performance or outcome ratings as scored by peers and by the teacher. In addition, students' performance as expressed by teacher marks did not correlate with interpersonal beliefs and conceptions of peer assessment. Only peer marks correlated positively ($r = .40, p < .01$) with value congruency. A high correlation was found between the marks for the product given in the peer assessment and those from the teacher assessment ($r = .86, p < .01$). In subsequent analyses perceived learning was used as a dependent variable to determine the impact of the assessment intervention in relation to interpersonal beliefs and assessment conceptions. Independent variables are measured as perceived by the participants as well.

To answer the first research question we compared pretest scores (i.e., at the beginning) and posttest scores on all variables measured, within the experimental group. Paired-sample t-tests revealed that at the end of the project (posttest) the students in the experimental group perceived value congruency as significantly higher, $t(33) = 2.24, p = .032$, Cohen's $d = 0.41$, than they did at the beginning of the project (pretest). Also, more trust in the peer as assessor was found, $t(34) = 2.32, p = .026$, Cohen's $d = 0.44$, than at the beginning of the project (pretest). In other words, the predicted changes in the interpersonal beliefs value congruency, and trust in the peer as assessor during the peer assessment intervention were confirmed.

Next, we examined whether scores on all variables within the peer assessment setting differed from those in the baseline condition of "no peer assessment intervention" (control group). The logistic regression analysis comparing the experimental and control groups (data from the measurement at the end of the project) showed that the overall model (with all variables entered) was significant at the .01 alpha level according to the model chi-square statistics. The model classified 78.9% of the students correctly (Nagelkerke $R^2 = .47$). This means that, for 78.9% of the students, the model predicted correctly whether they were in the experimental or the control group. These results indicate that it is possible to determine whether a student was part of the experimental (peer assessment) group or the control group on the basis of the scores on the interpersonal beliefs. In this model, value congru-

ency played a significant role ($B = 1.61$, $p = .016$) indicating that value congruency is a predictor of the group allocation. To analyse these results in depth a one-way analysis of variance was carried out. Results revealed a difference between the experimental and the control groups on two variables; specifically, psychological safety was higher in the experimental group, $F(1, 58) = 6.18$, $p = .016$, partial $\eta^2 = .10$, and value congruency was higher in the experimental group, $F(1, 57) = 11.91$, $p = .001$, partial $\eta^2 = .17$.

Table 2

Means, standard deviations, and correlations for the measures in the experimental group (N = 45)

Variable	M	SD	1	2	3	4	5	6	7	8	9	10
1. Psychological safety	2.06	0.54	--									
2. Trust in the self as assessor	2.23	0.58	.19	--								
3. Trust in the peer as assessor	2.32	0.57	.31*	.51**	--							
4. Value congruency	1.86	0.62	.51**	.18	.17	--						
5. Dependence of the self	2.20	0.82	.55**	.27	.52**	.40**	--					
6. Dependence of the peer	2.25	0.67	.44**	.45**	.61**	.43**	.66**	--				
7. Conceptions	2.41	0.75	.44**	.57**	.62**	.05	.35*	.38*	--			
8. Peer marks	7.42	1.56	.08	.26	.29	.07	.28	.40**	.25	--		
9. Teacher marks	6.87	2.30	.04	.25	.27	-.01	.22	.23	.28	.86**	--	
10. Perceived learning	2.78	1.04	.29	.43**	.23	.34*	.14	.23	.55**	.20	.17	--

* $p < .05$; ** $p < .01$.

Our second research question referred to the conceptual model of this study. We therefore analysed the possible mediating effect of conceptions of peer assessment on interpersonal beliefs and perceived learning, using hierarchical regression analysis (Rucker, Preacher, Tormala, & Petty, 2011). First, all independent variables were entered in order to predict conceptions. Results indicated that psychological safety ($\beta = .42$, $p < .005$), trust in the self ($\beta = .45$, $p < .001$), and trust in the peer ($\beta = .44$, $p < .005$) are significant predictors of conceptions of peer assessment. Additionally, value congruency predicts conceptions of peer assessment as well ($\beta = -.26$, $p < .05$). Dependence of the self ($\beta = -.18$, ns) and dependence of the peer ($\beta = -.04$, ns) were not found to predict conceptions of peer assessment. Second, the independent variables were entered in the analysis in order to predict perceived learning, first excluding, later including conceptions (Table 3). The results indicated a full mediation effect of conceptions of peer assessment regarding trust in the self as an assessor. Additionally, psychological safety and trust in the peer predict concep-

Chapter 4

tions, which in turn predict perceived learning.. Finally, value congruency is a positive predictor of conceptions, while it negatively predicts perceived learning.

Table 3
Hierarchical regression analysis predicting conceptions and perceived learning

	Conceptions			Perceived learning (Step 1: without conceptions)			Perceived learning (Step 2: including conceptions)		
	<i>B</i>	<i>SE B</i>	<i>B</i>	<i>B</i>	<i>SE B</i>	<i>B</i>	<i>B</i>	<i>SE B</i>	β
Psychological safety	.58	.19	.42**	.28	.36	.15	-.37	.37	-.19
Trust in the self as assessor	.60	.16	.45**	.76	.32	.41*	.10	.33	.06
Trust in the peer as assessor	.57	.18	.44**	.12	.37	.06	-.50	.36	-.27
Value congruency	-.31	.15	.26*	.47	.29	.28	.82	.28	.48**
Dependence of the self	-.17	.15	.04	-.20	.27	-.16	.01	.27	.00
Dependence of the peer	-.05	.18	-.18	-.12	.35	-.08	-.08	.31	-.05
Conceptions							1.09	.30	.77**
Adjusted R^2	.60			.17			.39		
ΔR^2							.22		
<i>F</i>	(6, 39) = 10.63, $p < .01$			(6, 40) = 2.40, $p < .05$			(7, 39) = 4.54, $p < .01$		
ΔF							(1, 32) = 13.02, $p < .01$		

* $p < .05$; ** $p < .01$.

4 Discussion

The answer to the question under which conditions peer assessment is successful is still inconclusive, despite a growing number of studies initiated by the increased interest in peer assessment as an assessment method to support learning (Van Gennip et al., 2009). Therefore, this study focused on students' perceptions of interpersonal beliefs as these relate to their conceptions of peer assessment and the perceived learning outcomes attained in a peer assessment setting. In order to better understand the nature of learning during peer assessment we investigated in particular the interpersonal beliefs influencing this process. Thus, this study aimed to contribute to a better understanding of students' involvement in the assessment as expressed by their perceptions of psychological safety, trust, value congruency, interdependence, and their conceptions of peer assessment, all related to perceived learning.

First of all, our results indicated that, in line with Hypothesis 1 students in the peer assessment group experienced at the end of the project more unanimity in goals, and more trust in the peer as assessor than at the beginning of the project. However, contrary to Hypothesis 1, scores on psychological safety, interdependence and trust in the self as an assessor did not differ between the two moments of measurement. Second, comparing experimental students' scores on the interpersonal beliefs at the end of the project with those from the control group revealed that psychological safety was higher in the experimental group as well as value congruency (i.e., there was more agreement between students). As it was expected (Hypothesis 1), these results showed that students in a peer assessment setting significantly feel safer and perceive more unanimity in goals than students in a traditional teacher-assessment setting. However, contrary to Hypothesis 1, trust and interdependence were not perceived differently by students from the experimental group, compared to control group students.

In answer to the second research question regarding the relations between the various variables, a hierarchical regression analysis showed that the relation between value congruency and perceived learning is fully mediated by the conceptions of peer assessment students hold. Furthermore, conceptions of peer assessment were predicted by psychological safety, and trust in the peer as an assessor, which in turn predict perceived learning. These findings seem to confirm Hypothesis 2 to a substantial degree, that is, interpersonal beliefs play a significant role in peer assessment settings. It was also found that conceptions of peer assessment act as a mediator between trust in the self as assessor and perceived learning from peer assessment. Contrary to Hypothesis 3, value congruency also appeared to be a direct predictor of perceived learning.

In relation to the role of value congruency, we have to acknowledge that value congruency as such had not been studied in peer assessment settings before. The findings, however, indicated that students in the peer assessment intervention achieve more unanimity on goals during the process, and experience higher value congruency than teacher-assessed students. Apparently, the peer assessment intervention resulted in more unanimity in goals, but contrary to our expectations we found that more unanimity in goals led to more negative conceptions of assessment (which was in turn positively related to perceived performance). However, the direct influence of value congruency is the other way round: the higher the degree of congruency (i.e., the more unanimity in goals) was, the higher students rated their learning gains. Findings from previous research on effective team learning (Jehn et al., 1999; Van Gennip et al., 2004) indicate that value congruency should be high in order for learning gains to increase. The present study confirmed these findings in a peer assessment setting when interpersonal beliefs were taken into account. The mediating effect of conceptions of peer assessment, however, is another matter. We may conclude that the process of peer assessment leads to

more agreement (high value congruency), but this does not mean that conceptions of peer assessment develop to more positive values.

The results do seem to indicate that psychological safety results in more positive conceptions of peer assessment, which in turn lead to a higher level of perceived learning. Previous research has already recognised the role of psychological safety in work environments (Edmondson, 1999) and our study indicates that it also influences learning in a peer assessment setting.

Following the literature on assessment quality (Falchikov & Goldfinch, 2000; Topping, 1998), the perceived quality of the assessor was measured in terms of trust in the peer as assessor. Trust in the peer as assessor, however, did not turn out to be a direct predictor of perceived learning. In contrast, trust in the self as an assessor appeared a predictor of perceived learning. Trust in the self and the peer as an assessor were both related to conceptions of peer assessment, which in turn affects perceived learning. The higher the degree of trust in the self and the peer as an assessor, the more positive students' conceptions of peer assessment are.

Contrary to Hypothesis 2 and 3 the hierarchical regression analysis showed no relation between interdependence (i.e., dependence on the self and the peer as assessors) and perceived learning, or between interdependence and conceptions of peer assessment. Unlike previous studies indicating that task interdependence leads to more learning in, for example, communication, helping, and information sharing (Crawford & Gordon, 1972), our study did not show an effect of interdependence on conceptions or perceived learning. Correlation analysis, however, did show significant correlations between interdependence subscales on the one hand, and interpersonal beliefs and conceptions of peer assessment on the other. This leads to the assumption that interdependence plays a significant role in the process of a peer assessment intervention and might lead to more alignment between the stakeholders involved, but this may not be directly related to the conceptions of peer assessment and perceived learning.

The present study explored peer assessment from a social perspective, acknowledging that interpersonal beliefs play a role in stimulating learning, that is, it explored specific relations among these interpersonal beliefs and their interrelations with conceptions of peer assessment and perceived learning. To date there have not been many studies investigating interactional processes in peer assessment (Strijbos, Ochoa, Sluijsmans, Segers, & Tillema, 2009). We explored the conceptual model of this study but future research will have to further validate the model, both in different peer assessment settings and in relation to the quality of assessment and students' and teachers' perceptions of the quality. Also of interest is the fact that in this study no correlation was found between perceived learning and the performance marks given by peers and teachers. A question that may be asked is whether performance measures are sensitive enough to capture the complexity of the learning that has been taking place. Therefore, in order to achieve a

more detailed picture of the differences between students, insights into the development of performance assessment (Kane & Mitchell, 1996) can point to ways in which to optimise the measurement of the learning effects of peer assessment in project-based classroom settings.

Furthermore, we would like to add that because all participants in this study were male, the generalisability of our results to populations including female students should also be examined. Finally, because of the small sample sizes, we suggest treating our results with caution. Future research should examine the generalisability of our results for larger sample sizes and in other educational settings, as well as in professional learning contexts.

To conclude, the present study stresses the importance of interpersonal beliefs in peer assessment, which feeds the need for further research on the social environment of peer assessment.

References

- Assessment Reform Group. (2006). *The role of teachers in the assessment for learning*. London: London University Institute of Education.
- Atwater, L. E., & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360° feedback. *Journal of Vocational Behavior*, *66*, 532-548.
- Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment and Evaluation in Higher Education*, *27*, 427-441.
- Bandura, A. (1986). *Social foundation of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Black, P. J., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, *5*, 7-74.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive-informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction*, *19*, 97-111.
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, *20*, 328-338.
- Crawford, J. L., & Gordon, A. H. (1972). Predecisional information-seeking and subsequent conformity in the social influence process. *Journal of Personality and Social Psychology*, *23*, 112-119.
- Crossman, J. (2004). Factors influencing the assessment perceptions of training teachers. *International Education Journal*, *5*, 582-590.
- Davies, P. (2002). Using student reflective self-assessment for awarding degree classifications. *Innovations in Education and Teaching International*, *39*, 307-319.
- Dochy, F. J. R. C., & McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation*, *23*, 279-298.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer-, and co-assessment in higher education: A review. *Studies in Higher Education*, *24*, 331-350.
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*, 350-383.
- Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International*, *32*, 175-187.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*, 287-322.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, *20*, 304-315.
- Hirschfeld, G. H. F., & Brown, G. T. L. (2009). Students' conceptions of assessment: Factorial and structural invariance of the SCoA across sex, age, and ethnicity. *European Journal of Psychological Assessment*, *25*, 30-38.
- Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. *Administrative Science Quarterly*, *44*, 741-763.
- Johnson, D. W. (1973). Communication in conflict situations: A critical review of the research. *International Journal of Group Tensions*, *3*, 46-67.
- Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Edina, MN: Interaction Book.
- Kane, M. B., & Mitchell, R. (1996). *Implementing performance assessment: Promises, problems, and challenges*. Mahwah, NJ: Erlbaum.

- Kennedy, K. J., Chan, J. K. S., Fok, P. K., & Yu, W. M. (2008). Forms of assessment and their potential for enhancing learning: Conceptual and cultural issues. *Educational Research for Policy and Practice*, 7, 197-207.
- Marton, F. (1981). Phenomenography – Describing conceptions of the world around us. *Instructional Science*, 10, 177-200.
- McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Teaching International*, 32, 302-313.
- Mesch, D., Marvin, L., Johnson, D. W., & Johnson, R. T. (1988). Impact of positive interdependence and academic group contingencies on achievement. *Journal of Social Psychology*, 128, 345-352.
- O'Connell, D. Q., & Dickinson, D. J. (1993). Student ratings of instruction as a function of testing conditions and perceptions of amount learned. *Journal of Research and Development in Education*, 27, 18-23.
- Orsmond, P., & Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assessment and Evaluation in Higher Education*, 21, 239-251.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academic Press.
- Pratt, D. D. (1992). Conceptions of teaching. *Adult Education Quarterly*, 42, 203-220.
- Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: current practices and new recommendations. *Social and Personality Psychology Compass*, 5/6, 359-371.
- Ryan, J., & Harrison, P. (1995). The relationship between individual instructional characteristics and the overall assessment of teaching effectiveness across different instructional contexts. *Research in Higher Education*, 36, 577-594.
- Searby, M., & Ewers, T. (1997). An evaluation of the use of peer assessment in higher education: A case study in the school of Music, Kingston University. *Assessment and Evaluation in Higher Education*, 22, 371-383.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29, 4-14.
- Sluijsmans, D. (2002). *Establishing learning effects with integrated peer assessment tasks*. *The Higher Education Academy*. Retrieved May 19, 2009, from http://www.heacademy.ac.uk/resources/detail/id437_establishing_learning_effects
- 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 and Evaluation in Higher Education*, 27, 443-454.
- Stanier, L. (1997). Peer assessment and group work as vehicles for student empowerment: A module evaluation. *Journal of Geography in Higher Education*, 21, 95-98.
- Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence-level in academic writing revision tasks: Are they critical for feedback perception and efficiency? *Learning and Instruction*, 20, 291-303.
- 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). Hershey, PA: IGI Global.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York: Macmillan.
- Tillema, H. H. (2009). Assessment for learning to teach: Appraisal of practice teaching lessons by mentors, supervisors, and student teachers. *Journal of Teacher Education*, 60, 155-167.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249-276.

Chapter 4

- 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* (pp. 55-87). Dordrecht, the Netherlands: Kluwer.
- Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education*, 25, 146-169.
- Van den Bossche, P., Gijsselaers, W., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments: Team learning beliefs and behaviors. *Small Group Research*, 37, 490-521.
- Van der Vegt, G., Emans, B., & Van de Vliert, E. (1998). Motivating effects of task and outcome interdependence in work teams. *Group and Organization Management*, 23, 124-143.
- 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.
- Van Gennip, N., Van den Bossche, P., Gijsselaers, W., & Segers, M. (2004, June). *Team learning: Dealing with diversity*. Paper presented at the 11th Educational Innovation in Business Education (EDINEB) conference, Maastricht, the Netherlands.
- Van Steendam, E., Rijlaarsdam, G., Serco, L., & Van den Bergh, H. (2010). The effect of instructional strategies on the quality of higher-order peer feedback in ESL/EFL. *Learning and Instruction*, 20, 316-327.
- Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20, 270-279.
- Wageman, R. (1995). Interdependence and group effectiveness. *Administrative Science Quarterly*, 40, 145-180.
- Winer, B. J. (1984). *Statistical principles of experimental design*. New York: McGraw Hill.