# Assessing together : Peer assessment from an interpersonal perspective

Gennip, A.E. van

# Assessing Together
# Peer Assessment from an Interpersonal Perspective

A.E. (Nanine) van Gennip

# Assessing Together
# Peer Assessment from an Interpersonal Perspective

**Proefschrift**

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. P.F. van der Heijden,
volgens besluit van het College van Promoties
te verdedigen op dinsdag 23 oktober 2012
klokke 13:45 uur

door

Nanine van Gennip
geboren te Hoorn (NH)
in 1979

Promotiecommissie:

*Promotores*
Prof. Dr. M. S. R. Segers
Prof. Dr. P. W. van den Broek

*Copromotor*
Dr. H. H. Tillema

*Overige leden*
Prof. Dr. W. H. Gijselaers
Dr. D. M. A. Sluijsmans
Prof. Dr. P. H. Vedder

# Dankwoord

*Een reis van duizend mijlen begint met een enkele stap (Lao Tzu)*

De reis van het vervolmaken van dit proefschrift is tot stand gekomen door het nemen van vele kleine en grotere stappen, die ik niet alleen had kunnen zetten. Om die reden wil ik op deze plek graag die mensen bedanken die mij in staat hebben gesteld deze reis te maken.

Het Koning Willem I College, en in het bijzonder Gérard Opstelten, Nico Stunnenberg, Wim van Nistelrooij, Rob Sars, Claus van der Heijden en Helmi van Zelst. Jullie hebben mij de mogelijkheden geboden binnen jullie afdeling een groot deel van mijn dataverzameling te doen, die aan de basis ligt van dit proefschrift. Maar dat niet alleen: jullie hebben die tijd onvergetelijk gemaakt door een ruime mate aan collegialiteit en plezier.

De Sint Antonius Academie (toen nog CAMERA), waar een groot deel van de dataverzameling op de werkplek heeft plaatsgevonden. Joke Wijntjes en Frans Klomp: bedankt voor het samenwerken op dit deel van mijn onderzoek.

Bernadette de Munk en Ronald te Loo. Tijdens een deel van deze reis werkte ik voor en met jullie bij Organise to Learn. Bedankt voor het bieden van ruimte en flexibiliteit waar nodig, ook als dit wel eens tijdens werktijd was. En mijn lieve collega's bij O2L: jullie hebben me laten zien wat samenwerken echt inhoudt, wat hebben we veel van elkaar geleerd.

Collega's van Rijn IJssel, bedankt voor het tonen van interesse tijdens het laatste deel van mijn proefschrift.

Ook dank ben ik verschuldigd aan studenten en studentassistenten van de afdeling onderwijsstudies in Leiden die op verschillende momenten tijd en energie in het onderzoek hebben gestoken. En natuurlijk mijn collega's van de afdeling onderwijsstudies, en in het bijzonder Marleen en Ilse: dank voor alle gezelligheid, bemoedigende woorden en vriendschap. Jan-Willem en Ron: als ik niet wist welke stap ik moest nemen was jullie behulpzaamheid onontbeerlijk. Michiel, wat is het ontzettend jammer dat je er niet meer bent.

Twee zeer waardevolle reisgenoten zijn Selma van der Haar en Karin Ploegh. Wat heerlijk om niet alleen de reis van het proefschrift, maar ook de reis van het leven met jullie te kunnen delen.

En natuurlijk Mien Segers en Harm Tillema. Gedurende de hele reis zijn jullie een rots in de branding geweest. Jullie inhoudelijke kennis was onmisbaar, maar juist en vooral onze interpersoonlijke samenwerking heb ik zeer op prijs gesteld. Ik

heb me door jullie altijd gesteund en begrepen gevoeld in de keuzes die ik heb gemaakt. Veel dank daarvoor.

Lieve ouders, dank voor de interesse in mijn werk, het bieden van een luisterend oor en het helpen volbrengen van deze reis. Lieve schoonouders, wat fijn dat jullie meereizen in ons leven, en voor de hulp die ook jullie hebben geboden. Liefste vrienden, jullie geven vele van mijn reizen kleur, ik zou niet zonder jullie kunnen. En natuurlijk liefste Christian, Imke en Maarten: wat is het fijn dat jullie er zijn! Onze gezamenlijke reis houdt hier niet op, en dat maakt me intens gelukkig.

Nanine van Gennip
Dieren, 2012

# Contents

# CHAPTER 1
# Introduction

A central theme nowadays is the responsibility people take for their own lives. The public as well as political discussion moves us more and more towards a society that relies on people's responsibility for their own life in general, and their environment and actions more specifically. This public shift to self-directedness brought about a reformulation of the concept of learning as well: Effective or meaningful learning now means that a learner constructs his own knowledge base that he or she can then use as a tool to interpret the world and solve complex problems with. This implies that learners have to be self-dependent and self-regulating, and have to be motivated to continually use and broaden their knowledge base. Finally, learners need meta-cognitive skills in order to reflect on their own and others' perspectives. Learners need to develop effective strategies to plan and monitor their own learning (Segers, Dochy, & Cascallar, 2003).

These changes in the current views on learning have led to the rethinking of the nature of assessment and argue for a balance between the call to account for what students learn (Assessment of Learning) and the need to create the classroom conditions under which they can and should learn (Assessment for Learning). During the past decades, many authors have tried to define the core features of the Assessment for Learning approach. They have consistently argued that assessment supports learning when teachers actively involve the learners in the assessment processes and when feedback is at the heart of the assessment process. In this respect, the use of peer assessment has been promoted. Although a variety of peer assessment arrangements are currently used in classrooms and peer assessment has gained increased attention by researchers, the learning effects of peer assessment are still not fully understood. More specifically, researchers have been referring to problems with the acceptance of peer feedback and the negative consequences of this lack of acceptance for students' learning from peer assessment. Peer assessment and as a crucial part of it, peer feedback, is an interpersonal process and it might be expected that the interpersonal context in which the peer

assessment takes place, is of utmost importance in its acceptance and thus its learning effect. Therefore, the learning effects of peer assessment might be better understood if we gain more insight into its interpersonal context. This is the core aim of this PhD dissertation.

## 1 Assessment for Learning

Closely linked to the assessment culture, many authors in the field of assessment in the past decade have argued the balancing of Assessment of Learning and Assessment for Learning. The Assessment for Learning perspective firmly positions assessment as an instrument for developmental purposes. The following definition of Assessment for Learning is provided by the Assessment Reform Group: "Assessment for Learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (Assessment Reform Group, 2002, pp. 1-2). Different authors have put forward a set of principles to describe what Assessment for Learning implies (e.g., Black & Wiliam, 1998; Assessment Reform Group, 2002; James & Pedder, 2006). In general, the following four principles are discerned:

1. Dialogue: If Assessment for Learning is to integrate assessment with the daily teaching and learning practice, it has to be a process of continual interaction between teachers and learners, a dialogue in which feedback provision and use are the core elements (Black & Wiliam, 2009).
2. Involvement of the learner: It is beneficial to their learning when learners are provided with the opportunities and the tools to assess to reach their learning goals themselves (Sadler, 1989). Self-assessments can be further complemented by facilitating interchanges through peer assessments, where students take on the role of assessors of each other's work (Black & Wiliam, 1998; Sadler, 1989; Topping, 1998).
3. Providing informative feedback: Feedback defines the learners' strengths and weaknesses and at the same time provides suggestions on the next steps to take in the learning process, whilst providing opportunities to improve upon the work.
4. Transparency of goals and criteria: Clarity on what needs to be achieved is a necessary part of the learning process. This transparency can be strengthened by involving the learners in the process of deciding on and formulating goals and criteria for assessing progress in their learning.

## 2 Peer assessment and the role of interpersonal beliefs

During the past decade, different models of assessment have been implemented that match the Assessment for Learning principles. In this respect, peer assessment has gained increased attention in research as well as daily classroom practice. The specific context of peer assessment as a tool to support learning is therefore the core object of study in this dissertation.

To date, peer assessment has been implemented in many classrooms. Based on a review of studies on peer assessment (e.g., Van Gennip, Segers, & Tillema, 2009) at least five different arrangements of peer assessment can be identified. They vary in the extent to which students are involved in the process from goal formulation to decision making. Firstly, in *peer marking* students score each others' work against a set of criteria, without further commenting on the work. Students are only involved in the scoring. *Peer feedback* takes this a step further, and allows students to comment on each others' work as well, often supplementing the scoring itself. *Peer grading*, thirdly, grants students the responsibility to make decisions about the outcomes of the assessment. However, feedback is not included in peer grading, while *peer appraisal* does include feedback. Finally, in *peer evaluation* students are not only involved in formulation of peer assessment criteria, scoring, giving feedback and the decision-making, but usually get to give input for the task formulation as well, at the start of the peer assessment process.

The aforementioned arrangements differ in the extent to which the students are involved in the different steps of the assessment process as well as in their role as assessor, from rater to feedback giver. It can be expected that when students are involved in the assessment process from the first steps on (goal formulation) and feedback is at the heart of the assessment process, their interpersonal beliefs will positively change and, in turn, acceptance of feedback and learning effects will increase. More precisely, when students participating in peer assessment collaboratively formulate learning objectives, this enhances similarity in opinion about what a team's task, goal or mission should be or value congruency (Jehn et al., 1999). Moreover, it makes clear to the students that they need each other's input in order to reach optimal learning effects (interdependence) (Johnson & Johnson, 1989; Mesch, Marvin, Johnson, & Johnson, 1988). In addition, by involving the learners collaboratively in the peer assessment process from the start and in this way increasing transparency, teachers enhance the students' feeling of psychological safety, a shared belief that it is safe to take interpersonal risks in a group of people (Edmondson, 1999). On the longer term, by gaining more experience with peer assessment, students' trust in both the self and the other as assessor will increase.

The influence of the aforementioned interpersonal beliefs (value congruency, interdependence, psychological safety, trust) on collaboratively sharing and build-

ing knowledge and in turn on team performance, has been evidenced in team learning research (e.g., Edmondson, 1999; Van den Bossche et al., 2006). When peer feedback is at the heart of the peer assessment process with peers sharing their insights on the others' work, it can be expected that, in accordance with the findings of team learning research, interpersonal beliefs will positively change and in turn increase learning gains.

## 3 Overview of the studies

**Chapter 2:**

**A review on the literature considering peer assessment from an interpersonal perspective.**[1]

This chapter reports a systematic literature review examining empirical studies on the effects of peer assessment for learning, focusing on the structural features of the assessment and the influence of interpersonal beliefs. The aim of this chapter is to provide a general idea of the relation between outcomes of peer assessment on the one hand and interpersonal beliefs and structural features on the other. A structural model of analysis is presented, which serves as the basis for this dissertation. The results of the review show that there is still little evidence of the effects of peer assessment on student learning and hardly any study addressed the interpersonal context in which a peer assessment intervention takes place. These results strengthen us in our conviction of the urgency of our empirical studies.

**Chapter 3:**

**Peer assessment fit for learning: framing the role of interpersonal beliefs.**[2]

This study is a first exploration of the role of interpersonal beliefs in a peer assessment setting, more precisely the beliefs of psychological safety and value congruency. The chapter aims to examine the influence of two different peer assessment arrangements on psychological safety and value congruency, with a teacher assessment arrangement as the baseline condition. The peer assessment and peer assessment$^{+}$ condition differ in the support students receive to develop beliefs of psychological safety and value congruency. The research questions are:

---

1 Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a team learning perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*, 41-54.

2 Van Gennip, N.A.E., Segers, M. S. R., & Tillema (2011, submitted). Arranging peer assessment: the role of interpersonal variables.

- Do students in a peer assessment condition hold more positive beliefs of psychological safety and value congruency than students in a teacher-based assessment condition?
- Do students who have been participating in a reflective session (peer assessment$^{+}$ condition) hold more positive interpersonal beliefs than students who did not participate (peer assessment condition)?
- In which stage of the peer assessment process do students in the peer assessment and peer assessment$^{+}$ conditions differ in interpersonal beliefs?

**Chapter 4:**

**Peer assessment as a collaborative learning activity.**[3]

The second empirical study of this dissertation attempts to get a more in-depth picture of the role of interpersonal beliefs in peer assessment. Firstly, it aims to confirm the results of the first study, in that a teacher assessment setting is compared with a peer assessment setting on how they affect the interpersonal beliefs of students. Second, we aim to explore the relation between interpersonal beliefs and perceived learning and the mediating role of students' conceptions. Previous research indicates the importance of conceptions of assessment for the acceptance and validity of assessments (e.g., Thompson, 1992; Pratt, 1992; Hirschfeld & Brown, 2009). Based on indications of Sluijsmans, Brand-Gruwel and van Merrienboer (2002) and Dochy, Segers and Sluijsmans (1999), we contend that, given that students in this study have no prior experience with peer assessment, the interpersonal beliefs they hold (psychological safety, value congruency, interdependence, trust) will influence how students conceive the peer assessment which in turn will influence perceived learning.

Hence, the first research question is: "Does participating in a peer assessment intervention result in a change in interpersonal beliefs (psychological safety, interdependence, value congruency and trust) and conceptions of peer assessment over time?" Pre-beliefs (prior to the peer assessment intervention) with post-beliefs (measured after the intervention) are compared, as well as a peer assessment condition with a control group. It is expected that a peer assessment intervention leads to higher scores on psychological safety, trust, interdependence, and value congruency (Hypothesis 1). Second research question is**: "**Is there a relation between (perceived) learning, interpersonal beliefs and conceptions?" It is 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 is fur-

[3] 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.

ther hypothesised that conceptions act as a mediating variable between interpersonal beliefs and perceived learning (Hypothesis 3).

**Chapter 5:**

**The role of interpersonal beliefs in 360 degree feedback.**[4]

The last study aims to study the role of interpersonal beliefs in a professional setting where peer feedback is the core of the assessment process and peer assessors have prior experiences with peer assessment. More specifically, in the organisations involved in this study, peer feedback is part of a 360 degree feedback system.

We question the role of participants' interpersonal beliefs (psychological safety, value congruency, interdependence) in their reactions to 360 degree feedback in terms of learning effects. Second, given that participants in this study have prior experiences with peer feedback and therefore have developed conceptions of peer assessment, we explore the role of participants' conceptions of peer assessment as a predictor (and not as a mediator as in chapter 4). Third, we include the role of trust, a variable which has formerly been evidenced as influencing participants' reactions to 360 degree feedback . It might be expected that interpersonal beliefs as well as previously developed conceptions of peer assessment predict the trust the participant has in himself and the peer as assessor. Trust might act as a mediator between the interpersonal beliefs the participants hold and their reactions to 360 degree feedback.

In this study (compared to chapter 4) we argue for a mediating role of trust. The argument lies in the definition of the construct trust. According to Edmondson (2002, p 7-8) "Trust, defined as the expectation that others' future actions will be favorable to one's interests, makes one willing to be vulnerable to those actions… The construct of trust, in contrast to psychological safety, pertains to anticipated consequences across a wide temporal range, including the relatively distant future." In the setting of this study, we ask participants to reflect on their experiences with peer feedback during the past years (and not on the peer assessment intervention they have just participated in as in chapter 4). Therefore, in this setting, we expect that participants beliefs of psychological safety, value diversity and interdependence, as well as how they conceive peer assessment and experience its transparency, have influenced on a longer term the trust they have in themselves and the other as raters and in turn how they react to 360 degree feedback in terms of learning benefits.

[4] Van Gennip, N., Gijbels, D., Segers, M., & Tillema, H. (2010). Reactions to 360° feedback: the role of trust and trust-related variables. International *Journal of Human Resources Development and Management, 10*, 362-379.

**Chapter 6:**
**General discussion.**

Finally, in chapter 6, we present a general discussion that contains an overview of the main findings and conclusions of the studies that were discussed in the previous chapters. We will discuss the results of the studies presented in this dissertation, and in which way they contribute to the understanding of the role of interpersonal beliefs in peer assessment, which has practically never been studied before. Additionally, theoretical and practical implications are discussed and directions for future research are suggested, as well as considerations for teachers implementing peer assessment as well as managers implementing 360 degree feedback. Furthermore, consequences on the quality of assessment are discussed and the limitations of the studies are outlined, as further ideas for future research.

## References


Assessment Reform Group (2002). *Assessment for learning: 10 principles*. Cambridge, University of Cambridge School of Education.

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*, 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability, 21,* 5–31.

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. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*, 350-381.

Edmonoosn, A.C. (2002). Managing the risk of learning: Psychological safety in work teams. Retrieved from http://www.hbs.edu/research/facpubs/workingpapers/papers2/0102/02-062.pdf

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.

James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal, 2*, 109-138.

Pratt, D. D. (1992). Conceptions of teaching. *Adult Education Quarterly, 42,* 203-220.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.

Segers, M., Dochy, F., & Cascallar, E. (2003). *Optimising new modes of assessment: In search of qualities and standards.* Dordrecht, Boston, London: Kluwer Academic Publishers.

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

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.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *66*, 249-276.

Van den Bossche, P., Gijselaers, 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 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.

# CHAPTER 2

# Peer Assessment for Learning from a Social Perspective: The Influence of Interpersonal Beliefs and Structural Features[5]

This paper reports a systematic literature review examining empirical studies on the effects of peer assessment for learning. Peer assessment is fundamentally a social process whose core activity is feedback given to and received from others, aimed at enhancing the performance of each individual group member and/or the group as a whole. This makes peer assessment an interpersonal and interactional process. Using this social perspective in order to study learning effects, we focus on the impact of the structural arrangement of peer assessment on learning, and the influence of interpersonal beliefs. The literature search, focusing on empirical studies measuring learning outcomes in a peer assessment setting, resulted in 15 studies conducted since 1990 dealing with effects (performance or perceived learning gains) of peer assessment. Our analysis reveals that, although peer assessment is a social process, interpersonal beliefs have hardly been studied; more specifically, they were measured in only 4 out of 15 studies. Moreover, they are not used to explain learning gains resulting from peer assessment. Finally, comparing the studies with respect to structural features reveals that, although the differences between the studies are significant, there seems to be no relation with the occurrence of learning benefits. The results of this review seem to indicate that research on peer assessment from a social perspective is still in its infancy and deserves more attention.

## 1 Introduction

Recent years have seen far-reaching developments with respect to the assessment of student learning. First of all, many studies (e.g., Black & Wiliam, 1998; Pellegrino, Chudowsky, & Glaser, 2001) indicate that formative assessment has a significant

[5] Based on: 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.

positive effect on student learning. These results have supported the growing attention in education for the implementation of assessment as a tool for learning. Secondly, inspired by social constructivism (which stresses students' responsibility for their own learning) we have paid attention to the role of students in assessments. Both developments have led to a search for adequate methods of assessment. One example of an assessment method in which students are playing an active role is peer assessment. This method is closely aligned with and embedded in the instructional process (Shepard, 2000). Peer assessment involves collaboration in the appraisal of learning outcomes by those involved in the learning process, i.e., students. Falchikov (1995) defines peer assessment in a clear and unambiguous way: "Peer assessment is the process through which groups or individuals rate their peers." Topping (1998) uses an even more explicit definition: "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). Vermetten, Daniels and Ruijs (2004) indicate that peer assessment is a direct appraisal not only of what has been learned (outcomes) but also of the where-to and the how of learning (processes).

The supposed beneficial effects of peer assessment are diverse. Peer assessment is said to enhance student learning (Davies, 2002). More specifically, using peer assessment helps students to develop certain skills in the areas of, for example, communication, self-evaluation, observation, and self-criticism (Dochy & McDowell, 1997). Literature reviews by Dochy, Segers and Sluijsmans (1999) and Topping (2003) indicate that although various studies seem to have found positive effects of peer assessment on learning, the results are still inconclusive. Moreover, it is unclear under what conditions peer assessment is effective. Dochy et al. (1999) and Falchikov (1995) refer to various problems that might arise given the social context of peer assessment, such as a lack of trust in the self and others as assessors, and friendship marking. Moreover, Dochy et al. (1999) indicate that the development of a shared understanding of the assessment procedures and criteria is a critical success factor in peer assessment. In sum, effective peer assessment requires attention to the social factors influencing the interactional process.

Therefore, the starting point for this literature review is that peer assessment is fundamentally an interpersonal process in which a performance grade exchange is being established and in which the core activity is feedback given to and received from others, aimed at enhancing the performance of an individual and/or a team or group as a whole. In this respect, this analysis of peer assessment studies is different from former review studies presenting the findings of peer assessment research in general (Dochy et al., 1999; Topping, 2003) or focusing on inter-rater agreement as one specific aspect of peer assessment (Falchikov & Goldfinch, 2000). Given the power of peer assessment as a tool for learning, and defining it as an inherently interpersonal process, we aim at gauging the influence of the social

nature of peer assessment on different learning benefits: achievement (as expressed in marks, grades, etc.), learning benefits as perceived by the students involved, and the beliefs students hold about peer assessment.

The present review aims to build on previous review studies on peer assessment (Dochy et al., 1999; Topping, 1998), focussing on effect studies. With respect to outcome measures a distinction is made between (1) learning benefits with respect to (increased) *performance and achievement*, (2) learning benefits as *perceptions* with respect to outcomes and (3) the *beliefs* students hold about peer assessment as a result of their experiences with this type of assessment. We conceptualize these beliefs as a collection of opinions and perceptions which are being influenced by the environment. The relevance of the latter outcome measure is argued by Crossman (2004). She refers to the role of prior assessment experiences, described as "student assessment histories" (p. 583), in students' approach to learning. These prior assessment experiences influence students' perceptions or beliefs on assessment, which in turn affect learning. Additionally, unlike former review studies, in this article we review peer assessment studies from an interpersonal perspective. We argue that interpersonal beliefs play a substantial role in the process of peer assessment, since these might interfere with the appraisal and affect relating to learning outcomes. Such a perspective involves factors that relate to group influences in action (Baron, 1994). Four interpersonal beliefs influencing learning from and with peers are discerned in this paper: psychological safety, value congruency, interdependence and trust (Van den Bossche, Gijselaers, Segers, & Kirschner, 2006). Moreover, in line with a large body of research on team learning (e.g., Dillenbourg, Baker, Blaye, & O'Malley, 1996; Webb & Palincsar, 1996) we focus on the structural features of peer assessment. This research has, for example, indicated the role of the heterogeneity of team composition in terms of disciplinary background and experience. These structural features may be expected to play a role in a peer assessment setting as well. For example, the choice for a face-to-face versus a distance peer assessment format might influence students' perceptions of psychological safety, and therefore directly and indirectly hinder or enhance the learning effects of the peer assessment. On the basis of Topping (1998), we address three clusters of structural features: (1) the description of peer assessment, (2) the interaction within peer assessment, and (3) the composition of the feedback group.

In sum, we reviewed studies addressing the effect of peer assessment on learning, taking into account the role of interpersonal beliefs as well as the extent to which structural features of the peer assessment format influence student learning.

### 1.1 Interpersonal beliefs in peer assessment

During the past decades research has focused on the role of the interactional factors constituting successful performance in group- and teamwork (Cohen & Bailey,

1997). It is the social context in terms of interactions that nourishes the willingness to engage in the (joint) effort to build and maintain mutually shared cognition (Barron, 2003; Crook, 1998). Barron (2003), for example, concluded from her multiple case-studies on sixth-grade triads that relational aspects of the interpersonal context need to be taken into account in order to understand what happens in learning groups. These groups have to deal with what Barron calls both a 'relational' and a 'content' space, which compete for limited attention. Her case study on less successful groups indicates that relational issues such as competitiveness and friendships can hinder or stimulate handling the insights constructed in the group. Several recent studies have suggested four interpersonal beliefs as particularly relevant: psychological safety, value congruency, interdependence, and trust (Edmondson, 1999; Lingard, Reznick, Espin, Regehr, & DeVito, 2002; Van den Bossche et al., 2006). There is evidence that the interplay between these beliefs influences the learning benefits of collaborative learning activities (such as peer assessment). These are discussed separately below.

### *1.1.1 Psychological safety*

Psychological safety can be described as a belief that it is safe to take interpersonal risks in a group of people (Edmondson, 1999). The idea that psychological safety may influence the learning effects of peer assessment has arisen because of the positive association of psychological safety with learning and group effectiveness found in several studies (e.g., Edmondson, 1999; Van den Bossche et al., 2006). Psychological safety, for example, prevents teams from perceiving differences in viewpoints as disagreements, creates room for framing a problem, and so promotes collaborative learning. As a result, psychological safety results in appropriate learning behavior and hence in better performance (Edmondson, 1999). Until now psychological safety has not been an explicit issue in peer assessment studies. Implicitly, however, it is acknowledged that peers have a tendency to make assessments on the basis of aspects such as friendship and uniformity (Dochy et al., 1999), so that psychological safety can be recognized as an influential factor in peer assessment. When peers perceive their environment as safe for interpersonal risk-taking they will be less prone to such conduct as, for example, friendship marking. Psychological safety, we contend, is a precondition for appraisal in a task-oriented and goal-directed way – a prime condition for assessment for learning as identified by the Assessment Reform Group (2006).

### *1.1.2 Value congruency*

Value congruency is defined as a similarity in opinion of what a team's task, goal or mission should be (Jehn, Northcraft, & Neale, 1999). Jehn et al. (1999) showed that value congruency should be high in order for teams to be effective. Van Gennip, Van den Bossche, Gijselaers and Segers (2004) also showed that work teams per-

formed better when value congruency was high. Integrating different perspectives and developing a shared understanding is crucial if teams are to perform well (Van den Bossche et al., 2006). The importance of developing a shared understanding is widely argued in reviews on peer assessment (Dochy et al., 1999; Falchikov & Goldfinch, 2000; Topping, 1998; Topping, 2003). The necessity of a common understanding is stressed especially with respect to assessment purposes, objectives, criteria and standards. For students involved in peer assessment the task of using their knowledge and skills to review, clarify, and evaluate the work of others is cognitively demanding. They are required not only to consider the objectives and purposes of the assessment task (Boud, 1995; Topping, Smith, Swanson, & Elliot, 2000), but also to contemplate the questions which assessment criteria to use, and which standards to employ in order to assess a piece of work as good or poor (Searby & Ewers, 1997). Because of the importance of generating assessment criteria and standards in order to enhance the learning effect of peer assessment, Boud (1995) and Ballantyne, Hughes and Mylonas (2002) recommend procedures to ensure that critical elements are included in the assessment criteria, and that criteria are amended whenever necessary to reach an optimal shared understanding between peers. Therefore, we contend that high value congruency will have a positive influence on peer assessment for learning.

### *1.1.3 Interdependence*

Interdependence between members of a group is a widely studied interpersonal factor in educational (e.g., Johnson & Johnson, 1989; Mesch, Marvin, Johnson, & Johnson, 1988) as well as organizational studies (e.g., Wageman, 1995) on team learning. A distinction can be made between outcome interdependence and task interdependence (Van der Vegt, Emans, & Van de Vliert, 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 in the sense that the performance of one specific piece of work depends on the completion of one or more other tasks (Van der Vegt et al., 1998). Studies have shown that task interdependence leads to more communication, mutual assistance, and information sharing than do individual tasks (Crawford & Gordon, 1972; Johnson, 1973).

When peer assessment is implemented as a tool to support learning, it is an integrated part of a collaborative learning process in which interdependence is “the glue that holds the members together” (Sluijsmans, 2002, p. 2). Peer assessment implies that multiple perspectives on reality are made explicit, and requires students to be individually responsible for an active contribution to group discussions. Learning from peer assessment occurs when there is a positive interdependence between the peers, i.e., when peers perceive that they are connected to each other

in such a way that the assessment task cannot be performed successfully unless everyone participates in a responsible manner. In this respect, Sluijsmans (2002) uses the concept of 'role interdependence', which "(…) occurs when the specific roles of assessor and assessee are assigned to the students. One student receives feedback from a peer and is then responsible in turn for giving feedback to another peer. In this situation a win-win relationship can be established" (p. 2). In this study, we have focused on task (role) interdependence.

#### *1.1.4 Trust*

Peer assessment for learning is commonly used to enhance students' shared responsibility for the assessment processes and learning. Improving students' responsibility for their own learning is a core argument for implementing new modes of assessment such as peer assessment (Birenbaum et al., 2006). In assessment literature it is argued that students who are actively involved in their learning as well as in the assessment process are more motivated, and therefore show more learning gains than passive students. However, several studies note that students feel uncomfortable criticizing one another's work, or find it difficult to rate their peers (Topping et al., 2000). This is partly a result of the 'novelty' of peer assessment in education. Generally, neither staff members nor students have much experience with this form of assessment. Ballantyne et al. (2002), citing various studies, indicate that it is common to find that students feel assessment to be the responsibility of teachers, who are recognized as experts. They conclude that students lack confidence in both their own and peers' abilities as assessors. For example, the results of Orsmond and Merry (1996) suggest that many students were skeptical about the added value of peer comments. McDowell (1995) indicates that students expressed concerns about their ability to provide constructive feedback and mark fairly. The influence of confidence or trust in both self and the other in relation to learning effects is hardly addressed in empirical studies. Therefore, Topping (1998) as well as Falchikov and Goldfinch (2000) suggest that future research should focus on the (perceived) quality of the peer assessor. In other words, appraisal of performance could possibly depend on the trust students have in their own and their peers' abilities as assessors, which is why we included trust in our review.

### 1.2 Structural features of peer assessment

Although researchers in general agree that peer assessment is a mode of evaluation in which peers appraise each other's learning (both process and outcomes), daily classroom practice reveals that peer assessment formats vary to a large extent. With respect to the organization of peer assessment, Topping (1998) presents a typology of seventeen different features. On the basis of the review model identi-

fied (Figure 1) and previous research (Van den Berg, Admiraal, & Pilot, 2006) we classified and grouped the seventeen features into three clusters (see Table 1). Cluster 1 is mainly a template for the description of peer assessment, and reflects the different ways to organize or arrange this assessment. Topping (1998) describes five basic questions regarding assessment features (parameters 1 – 10) (see Table 1): a) why, i.e., reasons for implementing peer assessment and purpose, b) what, i.e., objectives, curriculum areas, products/outcomes, c) when, i.e., time, d) where, i.e., place, and e) how, i.e., supplementary or substitutional; compulsory or voluntary; official weight.

Table 1
*Overview of clusters of assessment parameters*

| | Parameter | Range of variation |
|---|---|---|
| Cluster 1: Assessment description (why, what, when, where, how) | | |
| 1 | Curriculum area / subject (where) | All |
| 2 | Reasons for implementing peer assessment (why) | Of staff and/or students?<br>Time saving or cognitive/affective gains? |
| 3 | Purpose (why) | Summative or formative or both? |
| 4 | Objectives measured (what) | Examples: writing skills, presentation skills, professional skills |
| 5 | Outcomes (how) | Test score, open-ended feedback; quantitative or qualitative; credits, bonus point or other incentives for participation |
| 6 | Relation to staff assessment (how) | Substitutional of supplementary? |
| 7 | Official weight (how) | Contributing to assessees final official grade or not? |
| 8 | Place (where) | In/out class |
| 9 | Time (when) | Class time/free time/informally? |
| 10 | Requirement (how) | Compulsory or voluntary for assessors/assessees? |
| Cluster 2: Interaction | | |
| 11 | Directionality | One-way/reciprocal/mutual? |
| 12 | Privacy | Anonymous/confidential/public? |
| 13 | Contact | Distance or face to face? |
| Cluster 3: Composition feedback group | | |
| 14 | Year | Same or cross year of study? |
| 15 | Ability | Same or cross ability? |
| 16 | Constellation assessors | Individuals or pairs or groups? |
| 17 | Constellation assessed | Individuals or pairs or groups? |

The second cluster refers to the interactions within peer assessment, as results may vary according to who assesses whom (parameters 11– 13). This cluster covers directionality in peer assessment: one-way (from assessor to assessee), reciprocal (peers assess each other, e.g., in pairs), or mutual (all peers assess all peers). In

addition, peer assessment can differ as to level of privacy (anonymous, confidential, public) and contact between assessor and assessee (at a distance, or face to face).

The third cluster (parameters 14-17) refers to the composition of the feedback group. The assessor and assessee can differ in ability or have more or less the same level. In addition, the configuration of assessors and assessee can vary. One assessor to one assessee may be the standard constellation, but both assessors and assessees can be matched to individuals, pairs, or groups.

Given the interpersonal perspective of this review study, we analysed how, in the studies included in the review, structural features regarding the nature of peer interaction (cluster 2) and the composition of the peer group (cluster 3) are related to the learning effects of peer assessment.

## 2 Research questions

In our review of empirical studies into the effects of peer assessment on learning we intend to answer two main questions. Our structural model to review peer assessment studies is presented in Figure 1.

1. To what extent are the outcomes of peer assessment on learning (objective learning benefits, learning benefits as perceived by students, and beliefs) related to interpersonal beliefs (psychological safety, trust, congruency, and interdependence)?
2. To what extent are the outcomes of peer assessment on learning (objective learning benefits, learning benefits as perceived by students, and beliefs) related to structural features of the peer assessment format?

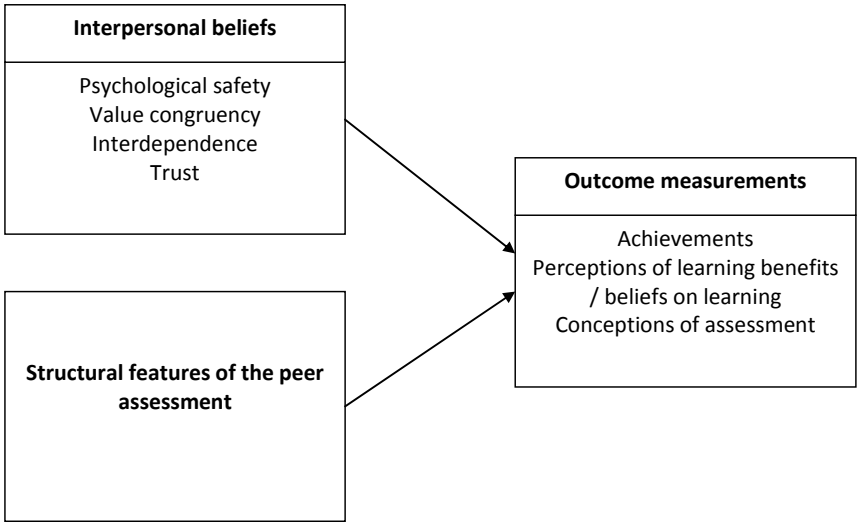


*Figure 1*
Structural model of analysis

# 3 Method

## 3.1 Selection of studies

In order to recover all relevant articles which evaluate peer assessment in relation to learning we conducted a literature search in the following databases: ERIC, PsychINFO, and EconLit. First, these databases were searched online. Following Falchikov and Goldfinch (2000), the search keywords were: peer assessment, peer grading, peer evaluation, and peer marking, each in combination with higher education, vocational education, adult education, professional education, and continuing education. No further pre-conditional criteria were added to the search. All available years of publication were included. This resulted in a total of 1275 articles. The range of these articles was very broad; peer evaluation, for instance, yielded many 'peer-evaluated' studies that were not related to peer assessment at all. Furthermore, this first selection included theoretical and review studies as well as empirical studies.

The next step was an analysis of abstracts. We preferred to comprehensively analyse the abstracts so as not to lose relevant articles by electronically supported analysis. For the comprehensive abstract analysis, inclusion criteria for the nature of the studies were: (1) the article or paper should describe empirical research; and (2) the assessment format should be peer assessment; and (3) the study should mention learning outcomes of some sort.

Abstracts of the 1275 hits were printed and their content was checked against all three inclusion criteria. If the information in the abstract was inconclusive the entire article was included for further analysis. Adopting the inclusion criteria led to a sharp drop in relevant articles: 83 article abstracts seemed to meet the criteria. In most cases, the reasons for not including a study were that peer assessment did not appear to be the predominant assessment format but was just one aspect of, for example, an article about assessment in general; or peer evaluation or peer review was described only as a constructive deliberation on a topic, without a formal final or intermediate judgement. Finally, it was clear from the abstracts that many articles did not describe empirical research.

Because peer assessment research has been implemented in a more systematic way since 1990, we decided to search for studies on peer assessment from 1990 till 2007. This resulted in the exclusion of 23 studies from further analysis. For the remaining 60 articles we collected full articles from Dutch libraries; when not available, the authors were contacted directly. Only one study (Phillips, 1992) could not be found, and could therefore not be included for further analysis. At this stage 59 articles were left which we regarded as relevant to our review.

As a third step, an in-depth full paper analysis was carried out, retaining the inclusion criteria used for the abstract analysis. This resulted in a sample of fifteen articles in all. The predominant reason for excluding articles at this stage was that full paper analysis revealed that learning benefits had not been measured as such. Finally, references in the full text of these articles were manually checked for other studies that possibly also answered our selection criteria. This procedure uncovered one additional article.

### 3.2 Method of analysis

We carried out a systematic literature review (Petticrew & Roberts, 2006). This form of review implies careful reading and analysis of separate articles on a given topic, and integrating the results of different studies into a coherent framework. The results were categorized on the basis of an analysis schema, depicted in Figure 1, which indicates the relations between the interpersonal beliefs, peer assessment format features, and various learning outcomes we identified. As a first step we constructed a descriptive table of the design and outcome variables found in the studies selected (Table 2). Subsequently, we used the categories of Figure 1 -- interpersonal beliefs and features of assessment -- in combination with the descriptive categories of research designs to construct Table 3 and 4, which list the studies according to interpersonal beliefs and beliefs, and according to structural features, contrasting these two aspects with learning outcomes.

### 3.3 Nature of empirical studies on peer assessment

The analysis of studies seems to indicate that until now the effect of peer assessment on learning has not been widely investigated. Our systematic search uncovered fifteen empirical articles written after 1990 that deal with learning benefits of peer assessment. Table 2 presents an overview of the nature of empirical studies on peer assessment.

Table 2 reveals that only three studies came close to an experimental control group design, controlling for peer assessment (Li & Steckelberg, 2004; Patri, 2002; Van den Berg et al., 2006). A study by Patri (2002) describes a bachelor's course in oral presentation skills. Students in both the experimental and the control group received a training session in which they could establish criteria by which to assess their peers' oral presentations. Finally, Van den Berg et al. (2006) implemented peer assessment in seven courses covering different types of writing assignments with a total of 168 students, of whom 37 were not in a peer assessment group.

Another three articles describe control group designs which differ as regards training in peer assessment. Lane and Potter (1998) divided their students into three groups. The first group did not receive any formal training or introduction to peer feedback. The second group was introduced to the idea of peer feedback, and extensively practised a peer assessment process. The third group received lectures and discussed the topic of peer assessment. McGroarty and Zhu (1997) divided their participants into two groups: an experimental group which received systematic training for peer assessment, and a control group which did not. They compared these groups on the basis of the "ability to critique peer writing, the quality of their writing and their attitudes toward peer revision and writing in general" (McGroarty & Zhu, 1997, p. 2).

The remaining nine studies were set up in a pre-test/post-test design, and administered questionnaires on two occasions in the peer assessment process. Horgan and Barnett (1991), for example, asked students to revise their papers on the basis of peer reviews by three reviewers. The quality of the papers before and after the peer review was compared as a performance measure.

With respect to the research design of the studies reviewed we noted that in fourteen of them the participants were university students or undergraduate students, with one exception (Lynch & Golen, 1992): here, lecturers' perceptions of the effects of peer assessment were described, not the perceptions of students.

Finally, the selected studies differ in terms of what is assessed. In seven studies written papers were assessed, and two studies used web-based projects. Other subjects of peer assessment were the quality of the lessons given by student teachers (Sluijsmans, Brand-Gruwel, & Merrienboer, 2002), oral presentations given by students (Patri, 2002), a poster (Orsmond & Merry, 1996) and the quality of feedback (McGroarty & Zhu, 1997). Regarding the effects found for peer

Table 2
*Overview of studies on peer assessment with respect to learning outcomes*

| Author | n | Design | Level | What is assessed? | Outcome: Performance (PF) or perceptions (PC) | Effect found: PF and PC | Outcome: Beliefs |
|---|---|---|---|---|---|---|---|
| Althauser, R., & Darnall, K. (2001) | 58 | Pre-test (unrevised essay)-post-test (revised essay) | University students | Written paper | PF: quality of revised essay assessed by tutor | +: quality of revised paper is higher when higher quality of peer review | - |
| Sluijsmans, D. M. A., Brand-Gruwel, S., & Merrienboer, J. G. (2002) | 93 | Pre-test/post-test control group design (control group: no training in peer assessment) | Higher education students | Quality of lesson | PF: three independent research assistants | +: higher quality of lesson in experimental group | overall vision on assessment |
| Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2002) | 57 | Pre-test (unrevised essay)-post-test (revised essay) | Undergraduate and senior high school students | Written essay | PC: perceived learning through PA | +/-: part of students perceived to have learnt | benefits; fairness |
| Patri, M. (2000) | 56 | Control group design (control group: no peer assessment) | University students | Oral presentations | PF: improvement in feedback quality | +: students are better assessors in experimental group | - |
| Purchase, H. C. (2000) | 96 + 123 | Questionnaires | University students | Interface design | PC: perceived learning through PA | +: perceived learning effect | - |
| McGroarty, M. E., & Zhu, W. (1997) | 169 | Pre-test-post-test control group design (control group: no training in peer assessment) | University students | Portfolio and quality of feedback | PF: tutor assessment of performance | +/-: experimental group showed no difference with control group | attitudes to PA |
| Orsmond, P., & Merry, S. (1996) | 78 | Questionnaires | Undergraduate students | Poster | PC: perceived learning through PA | +: perceived learning effect | benefits |
| Villamil, O. S., De Guerrero, M. C. M. (1998) | 14 | Pre-test (first draft) post-test (final draft) design | University students | Written paper | PF: amount of feedback suggestions carried out | +: most suggestions are carried out after PA | - |
| Lane, T., & Potter, B. (1998) | 53 | Pre-test (draft)/post-test (final paper) control group design (control group: no training) | University students | Written paper | PF: amount of feedback suggestions carried out | +/-: positive effect (more suggestions carried out) disappears after some time | - |
| Lynch, D. H., & Golen, S. (1992) | 78 | Questionnaires (filled in by lecturer) | Lecturers of business school | Nothing | PC: perceptions of tutors on performance progress of students | +: majority perceived PA as effective on performance and attitude | - |

| Author | n | Design | Level | What is assessed? | Outcome: Performance (PF) or perceptions (PC) | Effect found: PF and PC | Outcome: Beliefs |
|---|---|---|---|---|---|---|---|
| Horgan, D. D., & Barnett, L. (1991) | 74 | Pre-test (draft)-post-test (final paper) design | University students | Written paper | PF: quality of written paper assessed by tutor | +: better papers when peer suggestions were carried out | - |
| Stanier, L. (1997) | 36 | Questionnaires | Undergraduate students | Brochure for a specific sponsor/client | PC: perceived learning through PA | +: 94% of students reported a learning benefit of PA | benefits; feeling of comfort |
| Van den Berg, I., Admiraal, W., & Pilot, A. (2006) | 168 (131 exp. en 37 control) | Pre-test (draft)-post-test (final essay) control group design (control group: no peer assessment) | University students | Written paper | PF and PC: perceived learning through PA and quality of written paper assessed by tutor | +/-:Perceived learning benefit; PF: no difference in papers between control and experimental groups | - |
| Pope, N. K. L. (2005) | 160 | Two (gender) by four (group) experimental design | University students | Written essay | PF: quality of written paper assessed by tutor | +/-: no direct effects of PA on learning, but indirect through stress | - |
| Li, L., & Steckelberg, A. (2004) | 48 | Control group design (control group: no peer assessment) | University students | Web based project | PF and PC: perceived learning through PA and quality of project assessed by tutor | +/-: Perceived learning benefit; PF: no difference in quality of project between control and experimental groups | benefits; fairness |

assessment in general, a mixed picture emerges. Eleven studies found positive effects, i.e., these studies reported learning benefits as a result of peer assessment. One study reported no learning gains as a result of peer assessment, and two studies reported positive perceptions but no performance gains (see Table 2, last column). One study (Pope, 2005) reported indirect positive effects of peer assessment.

In sum, although we selected empirical studies that addressed the effects of peer assessment on learning, we could identify only three studies that adopted a controlled research design comparing groups of students with and without peer assessment. Moreover, empirical studies on the effect of peer assessment on learning seem to be restricted to higher education, although we explicitly included vocational education, adult education, professional education, and continuing education as search words.

In the next section an overview of the studies will be related to the research questions. First, articles taking interpersonal beliefs into account will be discussed. Next, format features will be discussed in relation to the effects of peer assessment. Finally, the learning benefits will be summarized as given in the articles included in this study.

# 4 Results

## 4.1 Role of interpersonal beliefs in peer assessment

In only four out of fifteen studies on peer assessment and learning benefits, interpersonal beliefs were studied in some form (see Table 3). Among the beliefs we identified in the studies value congruency and interdependence did not appear at all. Psychological safety was measured in Stanier (1997), and trust was described in three studies (Sluijsmans et al., 2002; Lin, Liu, & Yuan, 2002; Li & Steckelberg, 2004). To our surprise, none of the studies used these beliefs as an explanatory tool for learning benefits in the context of peer assessment: the beliefs were measured, but not explicitly related to learning benefits. In sum, it was not possible to derive a result regarding the influence of the interpersonal beliefs on peer assessment for learning.

### *4.1.1 Psychological safety*

Stanier's (1997) study applied peer assessment in a newly developed course for students of environmental sciences and geography. This course aimed to introduce students, in an interdisciplinary context, to studying and learning to collaborative work in groups. Students were assessed on their group work task at both group and individual level. Groups of students produced 'a brochure for a specific

Table 3
*Overview of interpersonal beliefs*

| Author | Features | | Interpersonal variables | | | |
|---|---|---|---|---|---|---|
| | Cluster 2 | Cluster 3 | Psychological safety | Value congruency | Inter-dependence | Trust |
| Althauser, R., & Darnall, K. (2001) | Mutual, confidential, distance | Same ability, assessors and assessed individual | - | - | - | - |
| Sluijsmans, D. M. A., Brand-Gruwel, S., & Merrienboer, J. G. (2002) | Mutual, ?, distance | same ability, assessors (individuals), assessed (groups) | - | - | - | +: trust in self as assessor |
| Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2002) | mutual, confidential, distance | Same ability, assessors (individuals), assessed (groups) | - | - | - | +: trust in other as assessor |
| Patri, M. (2000) | reciprocal, public, face to face | Same ability, assessors (individuals), assessed (groups of three) | - | - | - | - |
| Purchase, H. C. (2000) | mutual, public, face to face | Same ability, assessor (several individuals) and assessed (individual) | - | - | - | - |
| McGroarty, M. E., & Zhu, W. (1997) | mutual, public, face to face | Same ability, assessors (two individuals) and assessed (individual) | - | - | - | - |
| Orsmond, P., & Merry, S. (1996) | mutual, confidential, distance | same ability, assessors (individuals), assessed (pairs) | - | - | - | - |
| Villamil, O. S., & De Guerrero, M. C. M. (1998) | reciprocal, public, face to face | same ability, assessors and assessed in pairs | - | - | - | - |
| Lane, T., & Potter, B. (1998) | mutual, public, face to face | same ability, assessors and assessed individuals | - | - | - | - |
| Lynch, D. H., & Golen, S. (1992) | ? | ? | - | - | - | - |
| Horgan, D. D., & Barnett, L. (1991) | mutual, public, face to face | Same ability, assessors (three individuals), assessed (individual) | - | - | - | - |
| Stanier, L. (1997) | mutual, ?, ? | Same ability, assessors (individuals), assessed (groups) | +: safety in group work | - | - | - |
| Van den Berg, I., Admiraal, W., & Pilot, A. (2006) | varied | varied | - | - | - | - |
| Pope, N. K. L. (2005) | mutual, public, face to face | same ability, assessors (multiple individuals), assessed (individual) | - | - | - | - |
| Li, L., & Steckelberg, A. (2004) | mutual, confidential, distance | Same ability, assessors (two individuals) and assessed (individual) | - | - | - | +: trust in other as assessor |

sponsor/client aimed at a specific audience'. Criteria were formulated by the tutor(s) at an early stage, in consultation with students, and the brochure was assessed by peers (30%) as well as tutors (70%). The study addressed students' perceptions regarding the experience of peer assessment in general: did students perceive themselves to be empowered by the experience? Questionnaires included items about group work and peer assessment. Four items were included which resembled psychological safety: students reported that they enjoyed working in groups, there were not many personality clashes, they were working together on a task, and students perceived they were gaining by working with others. However, the relation between learning from peer assessment and the items resembling psychological safety was not examined. Finally, in terms of attitudes, the majority of students perceived peer assessment as an awareness-raising experience (74%), and said it made them think about the quality of other people's work (98%). However, 40% found peer assessment an uncomfortable experience as well (Stanier, 1997).

### *4.1.2 Trust*

Trust in the peer as an assessor was measured in the studies by Lin et al. (2002) and Li and Steckelberg (2004). In contrast, Sluijsmans et al. (2002) measured perceived trust in the self as assessor and their own assessment skills. In the study by Lin et al. (2002) senior high school students and undergraduate students were found to differ significantly in their opinions about being an assessor. High school students indicated that they did not feel that other peers had the knowledge required to evaluate their work. In contrast, undergraduate students were more neutral in their opinions. The students reported that they had 'benefited from marking peers' work', but no relation between learning effects and trust in the other as assessor was found. Beliefs about peer assessment were neither positive nor negative. Students were asked evaluative questions such as: 'It is worth to spend time on peer assessment' and 'Peers can assess fairly'. Students scored around the mean score of 3 (on a 5-point scale), which means that they were neither positive nor negative in their beliefs about peer assessment.

Li and Steckelberg (2004) randomly assigned 48 university students involved in a computer-based course entitled "Instructional Technology" to either an experimental group or a control group. Students had to develop a web-based project after studying the content area. In addition, the experimental group also had to judge their peers' performance, and received feedback from their peers so that they might improve their projects. Lin et al.'s (2002) questionnaire was re-used in this study, but applied in a different setting. Students in the experimental group (involved in peer assessment) thought that their peers did have adequate knowledge to evaluate their work (Li & Steckelberg, 2004), which resembles trust in the other as an assessor. Regarding learning benefits, results showed no significant

difference on project quality between the experimental group (involved in peer assessment) and the control group. Additionally, this study indicated that students were more positive about peer assessment than students in the 2002 study by Lin et al. Students reported that they had learnt more from peer assessment than from traditional teacher assessment; they considered peer assessment a worthwhile activity and felt they benefited from peers' comments (Li & Steckelberg, 2004).

Sluijsmans et al. (2002) implemented peer assessment in a course on designing creative lessons for student teachers: students (candidate teachers) assessed the quality of the lessons given by their peers. The experimental groups received an extensive training in peer assessment and practiced with peer assessment tasks, while control group students had extra time to study the domain content knowledge. Students in both the experimental and the control group had to write a qualitative peer assessment on their peers' lessons on creativity (craft work), which had been recorded on video. In both the experimental (receiving training in peer assessment) as well as the control group (no training in peer assessment) students seemed confident about their own assessment skills. However, the pre-test and post-test results indicate that this confidence did not increase after training in peer assessment skills. Regarding learning benefits, the results of this study reveal that training had a positive effect on the peer assessment skills themselves. More importantly, students from the experimental groups also performed better on the skill of designing creative lessons than students from the control group. This implies that training in peer assessment had a positive effect on students' performance in the content domain (Sluijsmans et al., 2002). Moreover, beliefs were more positive in the experimental group (involved in peer assessment): students felt more involved in the assessment than those in the control group (not involved in peer assessment), and the overall perception of assessment (e.g., 'I support the way I am assessed') grew more positive from pre-test to post-test.

## 4.2 Structural features in peer assessment

Table 4 shows an overview of how peer interaction and group composition as clusters (clusters 2 and 3) in the arrangement of peer assessment related to the different outcome measures.

### *4.2.1 The peer interaction cluster*

Peer interaction features of the peer assessment practice refer to how the feedback is organized: if it is given one-way, mutual, or reciprocal; if it is public or confidential, and face-to-face or at a distance. The analysis of the peer assessment studies included in this review indicated that feedback was organized in three different ways: a combination of (1) mutual, public and face to face feedback; (2) reciprocal, public and face to face feedback; or (3) mutual, confidential and distance feedback.

Table 4
*Features versus learning outcomes in peer assessment*

| | | Outcome measures Performance | Perceptions | Beliefs |
|---|---|---|---|---|
| **Cluster 2: Feedback provision** | Mutual, public, face to face | Lane & Potter (1998), Horgan & Barnett (1991), Pope (2005), McGroarty & Zhu (1997) | Purchase (2000) | McGroarty & Zhu (1997) |
| | Reciprocal, public, face to face | Patri (2000), Villamil & De Guerrero (1998) | | |
| | Mutual, confidential, distance | Sluijsmans et al. (2002), Li & Steckelberg (2004), Althauser & Darnall (2001) | Lin et al. (2002), Li & Steckelberg (2004), Orsmond & Merry (1996) | Sluijsmans et al. (2002), Lin et al. (2002), Orsmond & Merry (1996), Li & Steckelberg (2004) |
| | Varied/ Missing data | Lynch & Golen (1992), Van den Berg et al. (2006) | Stanier (1997), Van den Berg et al. (2006) | Stanier (1997) |
| **Cluster 3: Feedback provider** | Same ability, assessors (individuals), assessed (groups) | Sluijsmans et al. (2002), Patri (2000) | Stanier (1997), Lin et al. (2002), Orsmond & Merry (1996) | Sluijsmans et al. (2002), Lin et al. (2002), Orsmond & Merry (1996), Stanier (1997) |
| | Same ability, assessors (individuals), assessed (individuals) | Li & Steckelberg (2004), Althauser & Darnall (2001), Lane & Potter (1998), Pope (2005), McGroarty & Zhu (1997) | Li & Steckelberg (2004), Purchase (2000), Horgan & Barnett (1991) | McGroarty & Zhu (1997), Li & Steckelberg (2004) |
| | Other/ missing data | Villamil & De Guerrero (1998), Lynch & Golen (1992), Van den Berg et al. (2006) | Van den Berg et al. (2006) | |

When these three types of feedback are related to the outcome measure 'performance', a varied picture emerges: the reviewed articles are more or less equally divided over the three combinations of clusters. In other words, studies investigating student performance as a result of peer assessment differ in how they organize peer assessment. Moreover, there is no systematic distribution over clusters between studies measuring positive effects, and those reporting no effects of peer

assessment. For example, two studies with the same structural features show contrary results (McGroarty & Zhu, 1997; Lane & Potter, 1998). The studies varied in the amount of training in peer assessment given, but in both cases peer assessment procedure was mutual, public, and face-to-face. In the first study, students' drafts of written papers were reviewed by their peers, and students were allowed to use these reviews to revise their papers (McGroarty & Zhu, 1997). As hypothesised, the quality of revised papers was better than the quality of the drafts. Additionally, a quantitative analysis of the quality of the written feedback revealed that students in the experimental group, who had received training in peer assessment, outperformed students in the control group. However, there was no significant difference in the quality of revised papers between experimental and control groups. Additionally, regarding beliefs as outcome measure the attitudes toward peer revision were more positive in the experimental group. Aspects measured included items about the usefulness and meaningfulness of peer revision. Students reported peer assessment to be helpful (70%) and beneficial (72%). In the same line of research, and with the same study design, Lane and Potter (1998) revealed that students who had been introduced to the concept of peer assessment by means of an extensive training made the most changes per draft of their written paper. Lane and Potter (1998) argue that this indicates that it was easier to get used to the peer assessment process for these students than for students who did not receive any training. Unlike McGroarty and Zhu (1997), who measured students' performance at one point in time, Lane and Potter (1998) found that when students had to revise their papers a couple of times the difference between groups disappeared. This points to a training effect, implying that an effective training in peer assessment can help students to become comfortable with the peer assessment process, but also that through practice and familiarity the same level of comfort may be reached (Lane & Potter, 1998). However, the feeling of comfort was not measured.

Looking at the outcome measure 'perceptions of learning effects', we found that three out of six studies organized their feedback as mutually provided, confidential, and at a distance rather than face-to-face but (e.g., Lin et al., 2002; Li & Steckelberg, 2004; Orsmond & Merry, 1996). A case in point is the study by Althauser and Darnall (2001) (pre-test/post-test design), in which students peer-reviewed other students' written essays online (distance) in four assessment cycles, after which the students were able to revise their products. Results show that the better the written peer review, the higher the quality of the revised essay. Additionally, the quality of the peer reviews a student received was a significant factor for performance as well. This implies that students who receive high-quality peer feedback derive more learning benefits from peer assessment than those who receive low-quality feedback. Further, better-performing students produce better peer reviews (Althauser & Darnall, 2001). The study by Horgan and Barnett (1991) showed the same tendency, but with a different set of features (mutual, public,

face to face). In their study (pre-test/post-test design) about peer review of written work the results showed that better students were better reviewers, and better papers resulted from the acceptance of appropriate feedback. In other words, when students receive appropriate feedback from their peers by means of peer assessment, their learning gains in terms of the quality of the written papers are higher.

Finally, the same picture emerges when looking at studies measuring beliefs about the effects of peer assessment: most studies (66%) measuring beliefs apply mutual, confidential and distance feedback. For example, Orsmond and Merry (1996) focused on undergraduate science students studying Comparative Animal Physiology. Students (n = 78) worked in pairs (39 pair groups) on a scientific poster with the overall theme of neuropsychology. Individual students marked the posters of their peer groups (not their own). In terms of perceptions of learning benefits, 76 % of the students thought that peer assessment 'made them think more', while 69 % of the students perceived they 'learnt more with peer assessment'. Most students also found peer assessment 'helpful' and 'beneficial'. Further, Orsmond and Merry (1996) also measured perceptions of benefits as a result of peer assessment. They found, for example, that students believed that 'peer assessment makes you critical' (83%) and 'peer assessment makes you work in a structured way'.

### *4.2.2 The group composition cluster*

This cluster includes information about the feedback provider. Here, two combinations of features appear (see Table 4). In the first case, the assessors and assessees have similar ability, and assessors score as individuals (instead of groups or pairs), and assessees are individuals as well (e.g., Li & Steckelberg, 2004; Pope, 2005). In the other case the assessors and assessees are on the same level of ability, but the assessees are configured in groups or pairs instead of individually (e.g., Sluijsmans et al., 2002; Patri, 2002). These two combinations of features are more or less equally distributed over outcome measures. In other words, no relation can be found between the outcome measure applied on the one hand, and the combination of structural features in the peer assessment process on the other. An example of the first set-up (same ability, individual assessors, groups or pairs of assessees) is the study by Patri (2002), involving first-year bachelor students training oral presentation skills. Students in both the experimental and the control group received a training session in which they could establish assessment criteria for assessing their peers' presentations. Next, the students were divided into small groups in order to assess their peers. In the experimental group, students noted comments on the oral presentations. After each feedback session, individual group members completed peer assessment forms. The control group received no peer feedback. Results of the study by Patri (2002) revealed a significantly higher

agreement between peer assessment and teacher assessment in the experimental group than in the control group, which had not been involved in peer assessment. This seems to imply that students who engaged in a peer feedback session were better able to make judgements of their peers' oral presentations than the teacher (Patri, 2002), which was the performance measure of this study.

An example of a study in which both assessors and assessees are individuals is that by Purchase (2000). This study involved students in the technical domain who assessed each others' demonstration of an interface design in a Human-Computer Interaction course. The majority of students reported that "looking at other students' work is useful", and thus saw a perceived learning effect. On the other hand, the study by Pope (2005) on the effects of peer assessment on written papers of students involved in a research project showed no direct effects of peer assessment on learning. However, an indirect effect was demonstrated: when students were told they were to be marked by their peers, performance improved. This effect disappeared when stress was included in the analysis as a confounder (Pope, 2005). In other words, peer assessment seems to raise the stress level in students, and stress seems to enhance learning.

Two studies show different patterns of features. First, Villamil and De Guerrero (1998) implemented peer assessment involving peers with the same ability, and assessors as well as assessees operating in pairs. Their study (pre-test/post-test design) showed that peer review had a substantial effect on revision behavior: the majority of the revisions suggested were incorporated into final versions of papers. Second, Van den Berg et al. (2006) varied structural features. They developed an experimental design with different peer assessment settings. This study involved teachers and students from a history program. Again, peer assessment did produce positive learning outcomes in terms of students' perceptions: revised papers were perceived as better than draft papers, and students ascribed this improvement to the peer assessment process. However, results showed no significant differences in grades for the final products of the peer assessment groups as compared to non-peer assessment groups.

Finally, one article was purely descriptive, and therefore unsuitable to describe structural features of peer assessment (Lynch & Golen, 1992). In their study, instructors were asked to rate the effectiveness of peer evaluation on students' writing skills and their attitude toward writing. A small majority (54 %) of the instructors thought peer assessment to be 'effective' or 'very effective' for improving students' writing skills. A slightly larger percentage (59%) perceived peer assessment as effective for students' attitudes (Lynch & Golen, 1992).

To summarize, there are differences between the studies regarding such features of peer assessment as organization and characteristics of assessors and assessees. However, these differences were not related to the type of outcome measure. Additionally, no relation was found between studies reporting positive or

no effects of peer assessment, and the combination of assessment features chosen in the studies. We studied the influence of both interpersonal beliefs and structural features on learning, but did not find any study linking these three together.

### 4.3 Bringing interpersonal beliefs and structural features together

When we look at the structural features of the four studies describing interpersonal beliefs a pattern appears. Three out of four studies describe a peer assessment setting in which peers were of similar ability, assessors scored as individuals, and the assessees were groups of students (Stanier, 1997; Lin et al., 2002; Sluijsmans et al., 2002). In the interaction cluster it is more difficult to draw conclusions, because not all features were clearly described in the articles. Two studies described mutual, confidential and distance peer assessment formats (Li & Steckelberg, 2004; Lin et al., 2002), while Sluijsmans et al. (2002) described a mutual and distance process (confidentiality was not clear), and Stanier (1997) only states that the peer assessment was mutual.

## 5 Conclusion and Discussion

This literature review set out to disclose the effect of peer assessment on learning from a social perspective, unravelling the impact on learning benefits of both interpersonal and format features of the peer assessment setting. A first finding is that, to date, empirical evidence for the effect of peer assessment on learning is still scarce. Despite the fact that peer assessment was designed as assessment for learning with a formative purpose, research has not really concentrated on this component. According to our review, since 1990 only fifteen (refereed) empirical studies have been published. These studies differ in many aspects, which makes it difficult to draw conclusions with respect to effects of peer assessment on learning. First, only three of the fifteen studies have a control group design, with the experimental group implementing peer assessment and the control group using teacher assessment. Three other studies also use a control group design, but there the experimental and control groups differ in whether they received training in peer assessment or not. The remaining nine studies have a pre-test/post-test design, comparing students' learning gains before and after peer assessment. Studies inevitably also differed in quality and power. In this study, however, the reliability and validity of the experimental design, instruments and measures of the studies were not taken into account as such. It is clear that, in order to draw valid conclusions on the effect of peer assessment on learning, more high-power empirical studies using experimental or quasi-experimental approaches are needed. These studies should

take into account the structural features of peer assessment and interpersonal beliefs in relation to each other, and of influence on learning.

Second, there are large differences in the various operationalisations of the dependent variable 'learning', which again makes comparisons and generalisations difficult. Ten studies use performance measures regarding learning benefits of peer assessment (e.g., Sluijsmans et al., 2002; Villamil & de Guerrero, 1998) in terms of increased scores or performance. Six studies report perceived learning gains such as the study by Stanier (1997) in which perceptions of students on learning gains are described.

Third, although peer assessment is an inherently social activity, hardly any studies addressing the learning effects of peer assessment were conducted from a social perspective. Moreover, there is no single study relating interpersonal beliefs to the learning benefits of peer assessment. The few studies addressing interpersonal beliefs indicate that, in general, students were positive about two interpersonal aspects of peer assessment: students seem to feel safe when involved in peer assessment, and trust themselves and their peers as assessors. Only one study did not show positive perceptions about the specific belief of trust (Lin et al., 2002). There might be various reasons for the lack of research aimed at a better understanding of the effects of peer assessment from a social perspective. First, unlike peer assessment effect studies, most research on peer assessment has inter-rater agreement as its object of study. Implementing peer assessment as an alternative for or an addition to teacher assessment, most researchers are interested in the differences between peer and teacher marks (see the review study by Falchikov and Goldfinch, 2000). Second, the relatively small number of peer assessment studies focusing on learning effects of peer assessment primarily aim at finding empirical evidence for the benefits of peer assessment for student learning. Given that in many educational settings today peer assessment is still experienced by teachers and students as quite a revolutionary change in assessment practice, it is not surprising that the debate focuses more on evidencing the existence of effects than on understanding the conditions under which peer assessment can enhance student learning. Moreover, this focus on effects might have been strengthened by the fact that in many schools today assessment primarily serves summative and school accountability purposes. However, given the increased attention to assessment as a tool for learning, there is a growing need for research studies investigating the conditions under which peer assessment is beneficial for learning instead of investigating whether it works. Moreover, from the perspective that peer assessment is a powerful tool not only for evaluative decisions (marking, grading), but especially for the support of student learning, the social context in which this 'assessment for learning' takes place is a relevant object of study. In this respect, much information may be gleaned from the results of team-learning research evidencing the role of interpersonal beliefs if peers are to learn from each other. In short, there are

strong arguments for future research on peer assessment needs to take into account the role of the social nature of peer assessment in a more systematic way.

Regarding such features of intervention as peer interaction and group composition, no clear pattern emerged and research designs varied. Differences arose in, for example, face-to-face versus distance, 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 not possible to draw conclusions at this point, which makes further research necessary. Additionally, all studies analyzed in this article describe individual assessors as being of similar ability as their assessees. Heterogeneity in assessors' ability levels has so far never been investigated in relation to learning benefits of peer assessment. This, too, should be a subject for future research.

Given the nature of the studies analyzed we have opted for restraint in offering implications for educational practice. Because of the increasing implementation of peer assessment and the growing emphasis on assessment for learning, it is necessary to continue doing sound research on this topic. A perspective including social interactions and interpersonal beliefs is not a common way to look at assessment of student learning, and peer assessment in particular. However, future research should concentrate on a further investigation of the interactions between interpersonal beliefs and learning benefits of peer assessment, in order to unravel the processes that underlie the success (or failure) of new assessment forms. Finally, the structural features of peer assessment formats are already a recognized subject of research, but these have not been related to the learning effects of peer assessment. Therefore, research designs should be clear and well-grounded if any conclusions are to be drawn on the basis of these structural features.

## References


Althauser, R., & Darnall, K. (2001). Enhancing critical reading and writing through peer reviews: An exploration of assisted performance. *Teaching Sociology*, *29*, 23-35.

Assessment Reform Group (2006). *The role of teachers in the assessment of learning.* London: London University Institute of Education.

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education*, *27*, 427-441.

Baron, R. A. (1994). The physical environment of work settings. In M. Staw & L. Cummings (Eds.), *Research in organizational behavior* (Vol. 16, pp. 1-46). Greenwich: JAI press.

Barron, B. (2003). When smart groups fail. *The Journal of Learning Sciences*, *12*, 307-359.

Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., Wiesemes, R., & Nickmans, G. (2006). A learning integrated assessment system. *Educational Research Review*, *1*, 61-67.

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*, 7-74.

Boud, B. (1995). *Enhanced learning through self-assessment*. London: Kogan Page.

Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, *23*, 239-290.

Crawford, J. L., & Gordon, A. H. (1972). Predicisional information-seeking and subsequent conformity in the social influence process. *Journal of Personality and Social Psychology*, *23*, 112-119.

Crook, C. (1998). Children as computer users: The case of collaborative learning. *Computers & Education*, *30*, 237-247.

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.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189-211). Oxford: Elsevier.

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. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*, 350-381.

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.

Horgan, D. D., & Barnett, L. (1991, April). *Peer review: It works*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, USA.

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict, and performance in work groups. *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.

Lane, T., & Potter, B. (1998, March). *Teaching collaborative feedback strategies in intermediate writing*. Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages, Seattle, USA.

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*, 241-256.

Lingard, L., Reznick, R., Espin, S., Regehr, G., & DeVito, I. (2002). Team communications in the operating room: Talk patterns, sites of tension, implications for novices. *Academic Medicine*, *77*, 232-237.

Lynch, D. H., & Golen, S. (1992). Peer evaluation of writing in business communication classes. *Journal of Education for Business, September/October*, 44-48.

Mc Dowell, L. (1995). The impact of innovative assessment on student learning. *Innovation in Education and Training International*, *32*, 302-313.

McGroarty, M. E., & Zhu, W. (1997). Triangulation in classroom research: A study of peer revision. *Language Learning*, *47*, 1-43.

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.

Orsmond, P., & Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education*, *21*, 239-251.

Patri, M. (2002). The influence of peer feedback on self- and peer assessment of oral skills. *Language Testing*, *19*, 109-131.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academic Press.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences*. Malden, USA: Blackwell.

Phillips, J. (1992, March). *Writing in reading classes: The cliff-hanger effect*. Paper presented at the Annual International Rhetoric Council Conference, Hammond, USA.

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

Purchase, H. C. (2000). Learning about interface designs through peer assessment. *Assessment & Evaluation in Higher Education*, *25*, 341-352.

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 & 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 July 10, 2007, from http://www.heacademy.ac.uk/resources/ detail/id437_establishing_learning_effects

Sluijsmans, D. M. A., Brand-Gruwel, S., & Merrienboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment & 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.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *66*, 249-276.

Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy & E. Cascallar (Eds.), *Optimizing new modes of assessment: In search of qualities and standards*. Dordrecht / Boston / London: Kluwer Academic Publishers.

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*, 146-169.

Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education, 31*, 341-356.

Van den Bossche, P., Gijselaers, 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 & Organization Management*, *23*, 124-143.

Van Gennip, N., Van den Bossche, P., Gijselaers, W., & Segers, M. (2004, June). *Team learning: Dealing with diversity.* Paper presented at the 11th Educational Innovation in Business Education (EDiNEB) international conference, Maastricht, The Netherlands.

Vermetten, Y., Daniëls, J., & Ruijs, L. (2004*). Inzet van assessment: Waarom, wat, hoe, wanneer en door wie? Beslismodel voor een beargumenteerde keuze van assessmentvormen in onderwijs en opleiding* [Application of assessment: Why, what, how, when and by whom? Decision model for a substantiate choice of ways of assessment in education] (OTEC report 2001/13). Heerlen: Open University of the Netherlands.

Villamil, O. S., De Guerrero, M. C. M. (1998). Assessing the impact of peer revision on L2 writing. *Applied Linguistics*, *19*, 491-514.

Wageman, R. (1995). Interdependence and group effectiveness. *Administrative Science Quarterly*, *40*, 145-180.

Webb, N. M., & Palinscar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology*. New York: Simon & Schuster Macmillan.

# CHAPTER 3

# Arranging Peer Assessment: The Role of Interpersonal Beliefs[6]

The goal of the present study was to focus on the role of psychological safety and value congruency among peers and the ways in which it is affected by differentially arranged peer assessment conditions. We compare three conditions: (1) a teacher-based assessment condition; (2) a peer assessment condition, and (3) a peer assessment$^{+}$ condition, where the peer assessment and peer assessment$^{+}$ condition differ in the amount of involvement in peer assessment. Results indicate that teacher based condition differed significantly from both peer assessment conditions on psychological safety as well as value congruency.

## 1 Introduction

Peer assessment has become popular in school settings at different levels of education. It is described as "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" (Topping 1998, 250). This mode of assessment is being advocated as a strong tool to enhance learning (Dochy, Segers, & Sluijsmans, 1999; Falchikov, 1995; Sluijsmans, Brand-Gruwel, & Van Merriënboer, 2002; Van Gennip, Segers, & Tillema, 2009). As Tillema (2009) argues, peer assessment arranged as an exchange of appraisal information serves to scaffold each individual learner to (1) accept provided feedback, and (2) follow recommendations (Tillema, 2009). It is this informative scaffolding of further learning especially that makes peer assessment a powerful tool for the promotion of learning (James et al., 2006).

In addition to learning gains, multiple other benefits of peer assessment are claimed (Brown & Glasner, 1999; Lui & Carless, 2006; Falchikov, 1995; Pond, Ul-Haq, & Wade, 1995). It is argued that peer assessment increases student engage-

[6] Based on: Van Gennip, N.A.E., Segers, M. S. R., & Tillema (2011, submitted). Arranging peer assessment: the role of interpersonal variables.

ment in instruction and therefore encourages student autonomy (Sluijsmans et al., 2002). Moreover, it supports students in the development of critical judgment skills (Falchikov, 1995).

The definition of peer assessment as stated above, indicates that peer assessment is a collaborative activity between peers. Peer assessment is an activity taking place within the interpersonal context formed by the peer team. In their review study on team learning Akkerman et al. (2007) indicated that interpersonal beliefs or shared beliefs of the team characteristics emerge in groups from the interaction among the team members. Subsequently, it is shown that they form a context that stimulates or inhibits learning behavior. Given the interaction among peers differs according to the peer assessment arrangement they participate in, the main question to be dealt with is: How do students participating in peer assessment perceive the interpersonal context formed by their team and to what extent differ these interpersonal beliefs in different peer assessment arrangements?

Former research on peer assessment has shown some evidence that students' interpersonal beliefs in peer assessment settings differ to the extent to which students are involved in the different stages of the peer assessment process (Tillema et al., 2010). More concretely, within a peer assessment arrangement that has a low level of student involvement (e.g., peer marking in which students only give marks), students find it difficult to evaluate their friends and therefore show antagonism towards peer assessment (e.g., Hanrahan & Isaacs, 2001); they also perceive the grading of peers as risky and unfair (Kwan & Leung, 1996). Other formats of peer assessment arrangements (i.e, peer feedback or peer evaluation), however, deal more effectively with interpersonal relationships. More precisely, in the formats presented in these studies (Strijbos, Narciss, & Dünnebier, 2010) students are actively involved in different steps of the peer assessment process and in some cases receive training (Sluijsmans, Prins, & Martens, 2006) helping them understand the goals and criteria as well as practice the skills needed to assess peers. The results of these studies indicate that such peer assessment arrangements show an increased feeling of trust among the peers (Butler & Hodge, 2001; Keaten & Richardson, 1992; Pond & Ul-Haq, 1997). These findings indicate that the involvement of students in the different stages in a peer assessment process influences the degree to which students perceive the assessment setting as safe and free from interpersonal risk-taking. This finding is referred to as 'psychological safety' (Edmondson, 2002).

In addition, studies also indicate (e.g., Sluismans et al., 2002) that peer assessment arrangements in which students at the start of the peer assessment practice collaboratively define learning objectives result in an enhanced shared understanding of these learning objectives. According to Jehn, Northcraft and Neale (1999), this could be referred to as the degree to which students share values or 'value congruency'.

In sum, although evidence is still scarce, the aforementioned studies indicate that peer assessment arrangements which involve students in the early stages of

the peer assessment process positively affect interpersonal beliefs, which is an important condition for peers to learn from each other in a peer assessment setting. Based on these findings, in the present study, we focus on the role of psychological safety and value congruency among peers and the ways in which it is affected by differentially arranged peer assessments.

### 1.1 The interpersonal nature of peer assessment

Several authors (Falchikov, 1995; Marshall & Drummond, 2006) warned against the possible dangers or problems in ignoring the role of interpersonal relations in peer assessment when students are inexperienced with or involved only in parts of the peer assessment process. Topping (2003, p. 67) for example claims: "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." Studies of Dochy et al. (1999), Falchikov (1995), and Sluijsmans et al. (2002) refer to problems that arise in the arrangement of peer assessments. They mention students' hostility towards peer assessment when they first experience it, caused by a lack of trust in the self and the other as assessors, and point to friendship marking where peers mark their friends higher regardless of their performance (Dochy et al., 1999). These studies refer to the problem of psychological safety, a sense of confidence that the team will not embarrass, reject, or punish someone for speaking up (Edmondson, 1999, p. 354).

Moreover, several studies indicate (Sluijsmans, Dochy & Moerkerke, 1999; Rust Price & O'Donovan, 2003) that setting purpose and goals of assessments needs to involve students, since without explaining the rationale of an assessment it is hard to encompass its appraisal (Bloxham & West, 2007). These studies refer to the importance of value congruency for increasing learning benefits of participating in peer assessment.

#### *1.1.1 Psychological safety*

Although psychological safety has not been explicitly referred to in peer assessment studies, the aforementioned studies (e.g., Sluiijsmans et al., 2002; Dochy et al., 1999; Falchikov & Goldfinch, 2000) stress the importance of variables such as friendship and hostility and underpin the relevance of psychological safety. Peer assessment, as Edmondson states (2002, 3) "carries a risk for the individual of being seen as ignorant, incompetent, or perhaps just disruptive." There is a personal need to manage this risk to minimize harm to one's self-esteem especially in case one is being evaluated by another (Edmondson, 2002). Peer assessment arrangements, especially peer feedback and peer evaluation, can provide the conditions needed for low interpersonal risk-taking and therefore high psychological safety.

This is achieved, mainly, by involving students in the different steps of the peer assessment process. In this respect, Edmondson (2002) argues that seeking others' input and invite feedback and ideas from peers, which is de facto the core of a peer assessment practice, peers have a powerful positive effect on psychological safety. Moreover, she argues that organising a reflection meeting empowers those in lowered-status positions to speak up and to minimize the domineering tendencies of high-power individuals. We concur that when students are involved in the peer assessment arrangement - giving input, sharing ideas and giving feedback - this can contribute to students' safety in assessing peers and in turn support their learning. In addition, according to Edmondson (2002), the introduction of reflective sessions might affect the perception of psychological safety in a positive way.

### *1.1.2 Value congruency*

Value congruency refers to the degree to which peers agree on values about group processes and group tasks. The importance of a shared understanding between peers about the nature of assessment and the criteria on what and how to assess has been highlighted in many reviews on peer assessment (Dochy et al., 1999; Falchikov & Goldfinch, 2000; Topping, 1998; Topping, 2003). The study of Sluijsmans et al. (2002) confirms that practicing peer assessment enhances a shared understanding of the task. Reviewing, clarifying, and evaluating other persons' work is a cognitively demanding task for students which requires not only a common framework on objectives and standards in the assessment task (Boud, 1995; Topping, Smith, Swanson, & Elliot, 2000), but also the consideration (and acceptance) of multiple perspectives on assessing each others' work (Searby & Ewers, 1997). We assume that peer assessment arrangements which actively involve students foster greater value congruency. Moreover, a reflective session will support the feeling of value congruency as well.

## 1.2 Peer assessment arrangements

The many studies on peer assessment (Van Gennip et al., 2009) show high variety in the arrangement of peer assessment intervention and the degree of engagement or participation of students in the different parts of the assessment process (Birenbaum, 2007; Tillema et al., 2010). While in some cases students are only involved in the scoring or marking (Sivan, 2000; Topping et al., 2000) other arrangements include students as well in the criteria setting and feedback giving (Orsmond &Merry, 1996). Student involvement in and along the different stages of peer assessment adds to how they perceive the interpersonal context of peer assessment.

Based on an earlier review of studies on peer assessment (e.g., Van Gennip et al., 2009) at least five different arrangements of peer assessment can be identified in relevant literature. They vary in the extent to which students are involved in the

different stages of the process from goal formulation to decision making, including their role as assessor, from grader to feedback giver. Firstly, in *peer marking* students score each others' work against a set of criteria, without further commenting on the work. Students are only involved in the scoring. *Peer feedback* takes this a step further, and allows students to comment on each others' work as well, often supplementing the scoring itself. *Peer grading*, thirdly, grants students the responsibility to make decisions about the outcomes of the assessment. However, feedback is not included in peer grading, while *peer appraisal* does include feedback. Finally, in *peer evaluation* students are not only involved in formulation of peer assessment criteria, scoring, giving feedback and the decision-making, but usually get to give input for the task formulation as well, at the start of the peer assessment process. Figure 1 gives an overview of the different peer assessment arrangements.

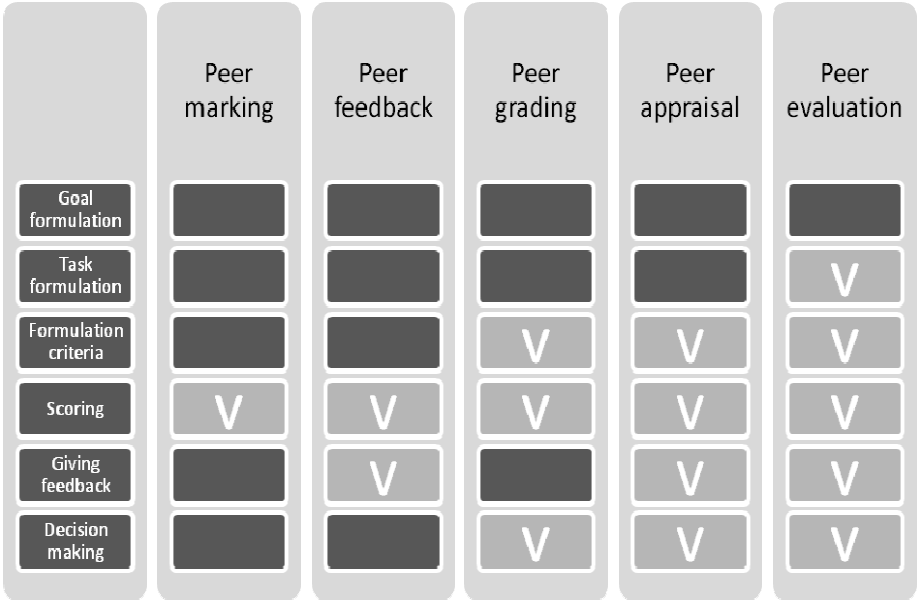


*Figure 1*
Overview of peer assessment arrangements

It was this variety in peer assessment conditions which challenged us to compare interpersonal beliefs not only between teacher assessment and peer assessment, but between different peer assessment arrangements as well. We therefore implemented a peer assessment arrangement with a reflective session, trying to create a more shared mindset between students using peer assessment, examining whether this leads to differences in perceptions of interpersonal beliefs (peer as-

sessment+ condition). Consequently, in this study, we focus on the relation between assessment arrangement and how students experience psychological safety and value congruency. In order to answer this question, we formulated the following research questions

1a Do students in a peer assessment condition hold more positive beliefs of psychological safety and value congruency than students in a teacher-based assessment condition?

1b Do students who have been participating in a reflective session (peer assessment+ condition) hold more positive interpersonal beliefs than students who did not participate (peer assessment condition)?

2 In which stage of the peer assessment process do students in the peer assessment and peer assessment+ conditions differ in interpersonal beliefs?

## 2 Method

### 2.1 Setting

Our peer assessment study took place in a large institute of vocational education in the Netherlands. 106 second-year male students participated in the study. Their ages ranged from 15 to 18. Students took a project-based course, teaching detailed technical skills in metal work and electronics. The assessment task was to design and construct a mechanical robot artefact: a moving device containing pneumatic and hydraulic elements.

We compare three conditions: (1) a teacher-based assessment condition; (2) a peer assessment condition, and (3) a peer assessment+ condition, where the peer assessment and peer assessment+ condition differ in the support students receive to develop beliefs of psychological safety and value congruency.

#### *2.1.1 Peer assessment condition*

This group consisted of 25 second-year students. During the project, groups received instruction in plenary sessions and worked on their artefacts as a group. At the start of the project students received a two-hour in-class instruction on the nature of peer assessment. The instruction included an overall explanation of the concept of peer assessment by the researcher. Additionally, fourteen appraisal criteria were formulated and students were instructed to use these in the appraisal of each other's work. At the end of the six-week project all groups presented their (robot) artefacts. After this presentation each group was assessed by their peers (i.e., not belonging to that group). Assessment was done on a special form that listed all eleven criteria, and students could rate their peers on a scale of 0 (poor) to 1 (good) for each of the criteria. The completed forms were collected after-

wards. The researcher calculated the average ratings and returned these to the project groups one week later.

### *2.1.2 Peer assessment+ condition*

Forty-five students participated in this condition. Peer assessment procedures in this condition were identical to those in the peer assessment condition. However, to enhance psychological safety and value congruency, a reflection meeting was organized where students reflected on and discussed among themselves how they approached the role of being an assessor and how they dealt with grading each other's work and hence, each other. This reflective session was implemented halfway down the course, in week 3 (out of a total of six weeks) and before the actual appraisal at the end of the course. At the start of the reflection meeting, the students completed a questionnaire with statements belonging to the scales 'psychological safety' and 'value congruency'. They received three cards with different colours: green, yellow and red. After the students had completed the questionnaires, the researcher read aloud the questionnaire statements one by one, after which every student showed one of the cards: red when the student disagreed, green when he or she agreed, and yellow in case of doubt. After this exercise, the researcher provoked a discussion among the students which he then moderated by posing questions like: 'Why do you agree/disagree?', 'Why does someone else thinks differently?', and 'How can you reach consensus in the group about this statement?' Meetings lasted approximately 1.5 hours, and all statements were discussed.

### *2.1.3 Teacher-based (or baseline) condition*

This is the regular appraisal condition for the course, lacking any kind of peer assessment. This group consisted of 36 second-year students. As a control group, these students received no training in peer assessment, but were assessed by the teacher only, on the same criteria the students in both peer assessment groups were instructed to use.

## 2.2 Research design

An experimental comparison of conditions was adopted for the peer assessment condition, the peer assessment+ condition and the teacher-based condition. For the teacher-based condition, end-of-course measurements are available. In both peer assessment conditions, prior measures (T0) as well as end-of-course measures (T2) of both interpersonal beliefs (psychological safety and value congruency) were included. Additionally, for both peer assessment conditions, psychological safety and value congruency were measured before the intervention (T1). Table 1 shows an overview of the research design of this study.

Table 1
*Overview of research design*

| | T0 | T1 | T2 |
|---|---|---|---|
| Teacher assessment baseline (n = 36) | | | Psychological safety<br>Value congruency |
| Peer assessment condition (n = 25) | Psychological safety<br>Value congruency | Psychological safety<br>Value congruency | Psychological safety<br>Value congruency |
| Peer assessment+ condition (n = 45) | Psychological safety<br>Value congruency | Psychological safety<br>Value congruency | Psychological safety<br>Value congruency |

## 2.3 Measurement instruments

All items were measured using 5-point Likert scales, and anchored by 1 (totally true) and 5 (totally untrue).

### *2.3.1 Psychological safety*

This scale measures the degree to which students perceive their group as safe for interpersonal risk-taking and was derived from Edmondson (1999). It consists of seven items. All items were measured using 5-point Likert scales, and anchored by 1 (totally true) and 5 (totally untrue). A sample item is: 'It is easy to ask my peers for help' (Cronbach's $\alpha$ = .86).

### *2.3.2 Value congruency*

The scale, adopted from a study by Jehn et al. (1999), measures value congruency as the differences that different group members perceive on group task and goal or mission. It consists of six items which were measured using 5-point Likert scales, and anchored by 1 (totally true) and 5 (totally untrue). Sample items here are: 'The group as a whole has one single goal' and 'Group members agree on what is important for the group' (Cronbach's $\alpha$ = .83).

## 2.4 Analysis

In order to answer research questions 1a and 1b, the three conditions (teacher-based assessment, peer assessment, peer assessment$^{+}$) were compared on the interpersonal beliefs (psychological safety and value congruency by means of ANOVA'S. For research question 2, in order to test at which stage in the peer assessment differences in interpersonal beliefs occur, multivariate analyses of variances were conducted (MANOVA).

# 3 Results

Means and standard deviations of variables measured in all groups are presented in Table 2.

Table 2
*Means and standard deviations*

| | Means (SD) | | |
|---|---|---|---|
| | Teacher assessment (n = 36) | Peer assessment (n = 25) | Peer assessment+ (n = 45) |
| Psychological safety | 2.40 (.68) | 2.79 (.42) | 2.95 (.54) |
| Value congruency | 2.51 (.78) | 2.88 (.61) | 3.14 (.62) |

Table 2 shows low mean scores (i.e., below 3 on the 5-point Likert scale) for both beliefs safety and congruency in all conditions. The variance in the teacher-based condition is found to be higher than in the peer assessment conditions.

## 3.1 Differences in interpersonal beliefs between conditions

To answer research questions 1a and 1b, we firstly investigated the differences between teacher assessment condition and peer assessment conditions: to what extent does teacher assessment differ from peer assessment, and more specific the various arrangements of peer assessment with regard to psychological safety and value congruency. Results show that conditions differed significantly on psychological safety $F(2, 99) = 9.11$, $p < .01$ as well as value congruency $F(2, 98) = 8.08$, $p < .01$. Further, results of Bonferroni corrected posthoc analyses show that, in the case of psychological safety, these differences appear between control condition and peer assessment condition (M = .40; $p < .05$), as well as between control condition and peer assessment$^{+}$ condition (M = .55; $p < .01$). In the case of value congruency, these differences only appear between control condition and peer assessment$^{+}$ condition (M = .62; $p < .01$).

## 3.2 In which stage of the peer assessment process do students in the peer assessment and peer assessment$^{+}$ conditions differ in interpersonal beliefs?

In answer to research question 2, a MANOVA of the two dependent measures was performed, which showed that there is an overall effect of time on psychological safety ($F(2,44) = 3.52$; $p = .034$; partial $\eta^2 = .074$) and value congruency ($F(2,44) = 5.43$; $p = .006$; partial $\eta^2 = .110$). Bonferroni corrected posthoc analyses show that there is a significant difference in value congruency between the start of the project and the intervention (M = .23; $p < .05$), but there is barely a difference between the intervention and the end of the project (M = .03; $p$ = ns). However, these post hoc analyses show no significant changes for psychological safety between the start

of the project and the intervention ( M = .14; *p* = ns), as well as between the intervention and the end of the project (M = .04; *p* = ns).

## 4 Discussion

The focus of this study is on whether interpersonal beliefs (psychological safety and value congruency) are differentially affected by different formats of peer assessment arrangement. Previous peer assessment research (Dochy et al., 1999; Tillema et al., 2010) points out the importance of a climate or assessment culture which is perceived as safe and in which participants agree on the goals and values of the assessment practice. In this study we focused on two interpersonal beliefs in particular: psychological safety and value congruency. First, we hypothesized that a difference between a peer assessment condition and a teacher-based assessment condition would occur in these two interpersonal beliefs. Second, we expected students in the peer assessment$^{+}$ condition would show higher levels of psychological safety and value congruency than students in the regular peer appraisal condition, exactly because they got a reflective intervention that raised awareness for the interpersonal beliefs that play a role in the assessment. Finally, we wanted to know where in the process of peer assessment differences in the interpersonal beliefs would occur.

In answer to the first research question, our results indicate that psychological safety is higher at the end of the project in the peer assessment conditions than in the teacher assessment condition. Value congruency is higher at the end of the project only in the peer assessment$^{+}$ condition compared to the teacher assessment condition: in the peer assessment$^{+}$ condition there is more unanimity of goals at the end of the project. This is in line with earlier research (Van Gennip, Segers, & Tillema, 2010) which showed that students in a peer assessment setting feel significantly safer and perceive more agreement in goals than students in a traditional teacher assessment setting.

In answer to the second research question, the time effect for psychological safety disappears after Bonferroni correction. Apparently, given the differences between the peer assessment conditions and the teacher-based condition, implementing peer assessment in itself is a powerful intervention in terms of increasing students' beliefs of psychological safety. For value congruency, differences seem to appear in the first half of the project; there is more unanimity in goals at the end of the project, which seems to take place in the first half of the project, before the intervention took place. This might be explained by the fact that discussion of goals and purposes with the students has been taken place before the reflection session. It seems that the reflection session has no value added to the stage of goal and purposes discussion in terms of increasing value congruency.

For future research, we suggest to measure the differential effects of all five peer assessment arrangements mentioned at the start of this study. They differ in terms of student involvement as well as the amount of feedback given which can be expected to influence the interpersonal context in which peer assessment takes place. Such research might advance our insights in how to organize peer assessment interventions in such a way that one can make optimal use of the advantages of students' interpersonal beliefs, and in addition to investigate the role of change in interpersonal beliefs on students' learning outcomes.

This study was conducted in a setting of secondary vocational education, where, given the educational setting and age of the students, peer pressure is a significant factor. It might be interesting to repeat our study in different educational settings to reveal the context-specificity of the role of interpersonal beliefs in peer assessment settings.

Moreover, due to the choice of studying interpersonal beliefs in real classroom peer assessment setting, the sample size in our study is rather small. Repeated studies are necessary to confirm our results.

Finally, although surveys are a valid method to detect relations between variables, quantitative research is necessary in order to have a deeper understanding of the meaning of the results. Observations of classroom behaviour of students, combined with the thinking aloud method to reveal students' motives for their behaviour, are interesting methodologies to pursue in future research.

Peer assessment has been introduced in different educational levels, although the practices differ in many aspects. Our study indicates that peer appraisal as a specific peer assessment arrangement is a powerful tool to enhance psychological safety and value congruency. Involving students in the different steps of the peer assessment process seems to be worthwhile to invest in.

## References


Akkerman, S., Van den Bossche, P., Admiraal, W., Gijselaers, W., Segers, M., Simons, R. J., & Kirschner, P. (2007). Reconsidering group cognition: From conceptual confusion to a boundary area between cognitive and socio-cultural perspectives? *Educational Research Review, 2*, 39-63.

Birenbaum, M. (2007). Evaluating the assessment: Sources of evidence for quality assurance. *Studies in Educational Evaluation, 33*, 29-49.

Bloxham, S. & West, A. (2007). Learning to write in higher education: students' perceptions of an intervention in developing understanding of assessment criteria. *Teaching in Higher Education, 12*, 77-89.

Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.

Brown, S., & Glasner, A. (Eds.) (1999). *Assessment matters in higher education: Choosing and using diverse approaches*. Buckingham: Society for Research into Higher Education & Open University Press.

Butler, S. A., & Hodge, S. R. (2001). Enhancing student trust through peer assessment in physical education. *Physical Educator, 58*, 30-42.

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.

Edmondson, A. (2002). Managing the risk of learning: Psychological safety in work teams. In West, M. (Ed.), *International Handbook of Organizational Teamwork*. London: Blackwell.

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.

Hanrahan, S., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research and Development, 20*, 53-70.

James, M., Black, P., McCormick, R., Pedder, D., & Wiliam D. (2006). Learning how to learn, in classrooms, schools and networks: Aims, design and analysis. *Research Papers in Education, 21*, 101-118.

Jehn, K. A., Northcraft, G. B. & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict, and performance in work groups. *Administrative Science Quarterly, 44*, 741-763.

Keaten, J. A., & Richardson, M. E. (1992). *A field investigation of peer assessment as part of the student group grading process.* Paper presented at the Western Speech Communication Association Convention, Albuquerque, NM.

Kwan, K., & Leung, R. (1996). Tutor versus peer group assessment of student performance in a simulation training exercise. *Assessment and Evaluation in Higher Education, 21*, 205-214.

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

Marshall, B., & Drummond, M. J. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education, 21*, 133-149.

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.

Pond, K., Ul-Haq, R. & Wade, W. (1995). Peer review: A precursor to peer assessment. *Innovations in Education and Training International, 32*, 314-323.

Pond, K., & Ul-Haq, R. (1997). Learning to assess students using peer review. *Studies in Educational Evaluation, 23*, 331-348.

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education, 28*, 147-164.

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.

Sivan, A. (2000). The implementation of peer assessment: An action research approach. *Assessment in Education: Principles, Policy & Practice, 7*, 193-213.

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

Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer-, and co-assessment. *Learning Environments Research, 1*, 293-319.

Sluijsmans, D. M. A., Prins, F. J., & Martens, R. L. (2006). The design of competency-based performance assessment in e-learning. *Learning Environments Research, 9*, 45-66.

Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*, 291-303.

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.

Tillema, H., Leenknecht, M., & Segers, M. (2010). Assessing assessment quality: criteria for quality assurance in design of (peer) assessment for learning - a review of research studies. *Studies in Educational Evaluation, 37*, 25-34.

Topping, K. (1998). Peer assessment between students in colleges and universities. Review of Educational Research, 68, 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.), Optimizing 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 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. 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.

# CHAPTER 4

# Peer Assessment as a Collaborative Learning Activity: The Role of Interpersonal Beliefs and Conceptions[7]

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, Sercu, & 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, Gijselaers, 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, Gijselaers, 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 feedback[8]), 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] 360º 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.

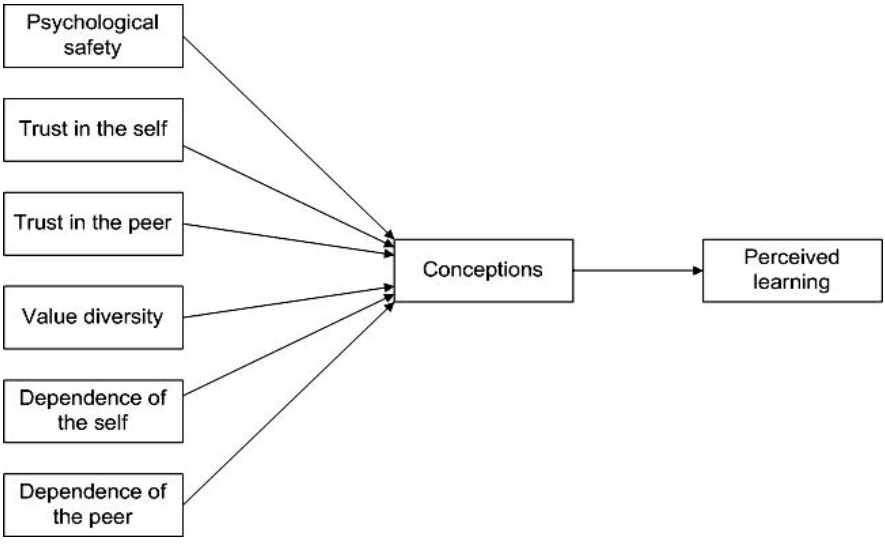


*Figure 1*
The conceptual model of the present study

# 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 (pretest), 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 α = .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 α = .64) as well as for Trust in the Peer as Assessor (Cronbach's α = .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 α = .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-

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) | | |
|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE B* | B | *B* | *SE B* | B | *B* | *SE B* | 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 | | |
| $\Delta R^2$ | | | | | | | .22 | | |
| *F* | (6, 39) = 10.63, $p < .01$ | | | (6. 40) = 2.40, $p < .05$ | | | (7, 39) = 4.54, $p < .01$ | | |
| Δ*F* | | | | | | | (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 leaded 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., & Wiliam, 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, fromhttp://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.

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., Gijselaers, 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., Gijselaers, 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., Sercu, 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 assesment 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.

# CHAPTER 5

# Reactions to 360 Degree Feedback: The Role of Trust and Trust-related Variables[9]

One of the most significant changes within the workplace over the past decades has been the growing emphasis on the development of competencies rather than specific job skills. This change has resulted in an increased use of multiple-rater feedback systems such as 360 degree feedback, which is the focus of this article.
The aim of this study is to provide insight into the factors that contribute to employees' perceived reactions to 360 degree feedback. In this study we focus specifically on a developmental use of the assessment instrument, arguing that trust mediates interpersonal beliefs, such as psychological safety, value congruency, and interdependence when we consider the outcome of a developmental assessment. Furthermore, conceptions of assessment and the transparency of the process of 360 degree feedback are expected to affect employees' reactions to assessment.
Results reveal that trust in the other as assessor partly mediates the relation between value congruency and reactions to 360 degree feedback, and as such is a significant predictor of reactions to 360 degree feedback. Conceptions and transparency are predictors of trust in the self, which does not predict reactions to 360 degree feedback. Finally, the interpersonal beliefs of psychological safety and interdependence have an indirect effect on reactions to 360 degree feedback: they are predictors of trust in the other, which in turn predicts reactions to 360 degree feedback ..

## 1 Introduction

The fast changes in occupational structures and in work content and organisation have challenged companies to develop new ways to ensure that the competence level of the workforce meets the demands of the changing workplace (Tynjälä, 2006). From a Human Resources perspective the development of individual employees has increased in importance, and is seen as indispensable in the learning organisation. New ways of assessment have therefore been developed which focus

[9] Based on Van Gennip, N., Gijbels, D., Segers, M., & Tillema, H. (2010). Reactions to 360° feedback: the role of trust and trust-related variables. *International Journal of Human Resources Development and Management, 10*, 362-379.

on employee development rather than employee selection (Lievens et al., 2003). The emphasis on the need for a flexible workforce and for competence development has increased the use of multiple-source multiple-rater (MSMR) feedback systems such as 360 degree feedback (Fletcher, 2001). This type of feedback can be described as involving multiple raters, often including the participants themselves, in the assessment of individuals. More specifically, 360 degree feedback includes feedback solicited from 'significant others', using a standardised assessment instrument (Thornow, 1993). These significant others typically include colleagues and peers as well as subordinate employees, managers and customers (Tillema, 2001). Therefore, as a multi-rater instrument 360 degree feedback provides informative assessments that involve reciprocal learning partnerships, introducing multiple perspectives from different sources. As an assessment tool multi-rater instruments can inform the learner about different performance aspects as well as progress in competence from multiple perspectives, and therefore is a powerful developmental tool for professionals' learning.

There are indications that in 360 degree feedback psychological safety, value congruency, and interdependence play a role. However, empirical evidence is to a large extent still lacking. Moreover, prior research on assessment involving multiple raters, mostly conducted in school settings (e,g., peer assessment research) indicate that transparency of the assessment process (Sluijsmans et al., 2002), as well as the way students perceive it (Hirschfield & Brown, 2009) affect reactions to the assessment. In short, we studied the influence of 1) trust; 2) the perceptions of interpersonal beliefs, including a) psychological safety, b) value congruency, c) interdependence); 3) transparency of the feedback system, and 4) the conceptions of 360 degree feedback. In section 3 we will discuss these variables in more detail.

## 2 Feedback from 360 degree instruments

In recent years, 360 degree feedback systems have received increased attention as developmental HR tool. The aim is "to provide constructive feedback in a climate in which one's growth is fostered and there is room for improving one's weaknesses without immediate negative consequences" (Van der Heijden & Nijhof, 2004, p. 494). 360 degree feedback is considered as a relevant instrument by which to improve employees' performance (Atwater & Brett, 2005). However, the results of the Smither, London and Reilly (2005) meta-analysis indicate that effect sizes of multi-rater feedback are rather small. This result is in line with the pivotal review study by Kluger and DeNisi (1996) on the effects of feedback on performance, showing that feedback does not always result in performance improvement (e.g., Thornow, 1992). In their study over one third of cases even showed a decrease in performance after the feedback intervention. As Smither et al. (2005) conclude,

these results indicate the need for research that focuses on the conditions under which multi-source feedback is beneficial, instead of investigating whether it works at all. Studies in the area of 360 degree feedback (e.g., Atwater & Brett, 2005; London & Smither, 1995) point to the importance of people's reactions to feedback in terms of cognitive and emotional evaluations, because these have 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). Therefore, in our study the reactions to 360 degree feedback were conceptualised (following Atwater & Brett, 2005) as a perceived improvement in functioning as a result of the feedback, i.e., the perceptions of the relevance of feedback received to workplace learning. Hence, the aim of the present study is to provide insight into the factors that contribute to employees' perceived reactions to 360 degree feedback.

In this respect, recent studies on 360 degree feedback have focused on the role of the characteristics of the feedback (such as sign and source) and individual dispositions (such as self-efficacy, trust, emotional stability, openness to experiences, and conceptions of feedback) (Atwater & Brett, 2005; Baily & Austin, 2006; Becton & Schraeder, 2004). A relevant case is discussed by Atwater and Brett (2005), who address feedback characteristics (positive or negative; self-other discrepancies) in relation to reactions to feedback, and managers' engagement in follow-up activities. Additionally, they investigated the role of individual dispositions: trust, self-efficacy, emotional stability, openness to experiences, and conceptions of feedback, as well as the influence of the source of the feedback (direct reports, peers, or managers). The results indicate the relevance of the sign (positive or negative) as well as the participants' conceptions of 360 degree feedback: there were more positive reactions after positive than after negative feedback, and a more positive attitude towards using feedback resulted in more motivated employees afterwards.

In the Atwater and Brett (2005) study there are no indications for the influence of the individual dispositions of trust, emotional stability, and openness. The study by Baily and Austin (2006) confirms the Atwater and Brett (2005) results with respect to the influence of a favourable feedback on subsequent performance. Moreover, their study indicates the role of two individual dispositions in the relation between feedback and performance: initial self-assessment (before feedback was given), and self-efficacy before participation in the feedback process. Finally, in both the Atwater and Brett (2005) and the Baily and Austin (2006) studies the source of the feedback seems to influence participants' reactions to the feedback received. Baily and Austin (2006) conclude: "Further research is needed that examines the credibility of different rater sources and factors influencing focal individuals' attentiveness to particular sources" (p. 63). This conclusion is supported by

Becton and Schraeder in their descriptive article (2004, p. 26): "The importance of the credibility of the feedback source cannot be underemphasised." Becton and Schraeder (2004) point to the lack of research on trust in the assessor as a predictor of reactions to multi-rater feedback.

Therefore, in this study we investigated the role of trust on the reactions of 360 degree feedback, which in our idea is not only trust in the other, but also trust in the self as an assessor. Additionally, we wanted to understand which factors influence these two types of trust thus indirectly affect reactions to 360 degree feedback. We believe that acceptance of feedback, especially in the case of a multi-rater, i.e., interpersonal assessment tool to a large extent depends on factors relating to interpersonal dispositions. Or, as Van der Heijden and Nijhof (2004, p. 494) describe: "A fruitful application of 360 degree appraisal depends upon a climate in which people can inform one another of strengths and weaknesses in performance at a particular career stage." It is clear that multi-rater assessments make interpersonal contact necessary, and therefore is inevitable to acknowledge the influence of interpersonal beliefs and trust in 360 degree feedback settings. Considering the lack of research on this aspect so far, it would be of interest to determine how interpersonal beliefs come into play in a multi-rater environment. Further, earlier research indicates that employees' conceptions of 360 degree feedback play a role as well (Atwater & Brett, 2005): the more favorable these are, the higher the degree of trust in the self as an assessor. Additionally, based on the results of empirical research in the field of classroom assessment and inspired by the arguments presented in studies on 360 degree feedback, we pose that conceptions affect the acceptance of and reactions to feedback from 360 degree instruments. Finally, there are clear indications in classroom assessment studies that the transparency of the assessment process influences student reactions to assessment. Therefore, it can be expected that the trust employees have in themselves and in others as assessors is influenced by how transparent they perceive the assessment practice. We assumed that perceived transparency of the assessment process could correlate with trust since it might build confidence in the outcome of the assessment.

## 3 The present study

The aim of our investigation was to take a closer look at the influence of interpersonal beliefs on the reactions to 360 degree feedback. We have conceptualised these reactions as perceived improvement in functioning as a result of 360 degree feedback. In order for 360 degree feedback to be accepted, trust (as 'having confidence in the appraisal given') is a crucial factor (Arnold, 2004). However, assuming an influence of trust on reactions to 360 degree feedback raises the question how exactly a high level of trust is achieved. In this research, we analyzed and tested the

assumption that *interpersonal beliefs* to a large extent determine the degree of trust in the assessor. Conceptions and transparency are also expected to predict trust in a 360 degree feedback setting. We studied these factors as influencing trust in a performance-oriented environment. This assumption acknowledges the personal relations between raters, which may impinge on the confidence with which their ratings are accepted by assessees. More specifically, the interpersonal beliefs, conceptions, and transparency have a possible effect on trust, which in turn will affect the outcome measure of perceived learning. Therefore, we regard trust as a mediator in the relationship between interpersonal beliefs, conceptions, and transparency on the one hand, and employees' reactions to the 360 degree feedback they receive.

In short, all independent variables, as well as trust in the self and trust in the others, are expected to influence reactions to 360 degree feedback. Additionally, trust in the self and trust in the other are expected to mediate between the independent beliefs and reactions to 360 degree feedback. Our model of variables influencing the acceptance of information from 360 degree feedback is displayed in Figure 2; we will discuss each variable in more detail.

### 3.1 Trust

Taking control of their learning process motivates people, and 360 degree feedback offers an opportunity to be more involved in their own learning and development. However, because of the unconventional combination of (mostly inexperienced) assessors, new challenges appear regarding the trust people have in themselves and others as assessors. For example, in her studies on peer assessment McDowell (1995) has indicated that participants expressed concerns about their ability to provide constructive feedback and assess fairly. In addition, the general objectivity or fairness of assessments in which peers are assessors is sometimes doubted by participants (Sluijsmans et al., 2002). Ballantyne, Hughes and Mylonas (2002) even conclude that students lack confidence in both their own and peers' abilities as assessors. Because of the interpersonal component in assessment settings such as 360 degree feedback and peer assessment it is all the more striking that the role of trust has hardly been examined so far. Taking into account the lack of research on trust as a predictor of reactions to feedback, which is indicated by Becton and Schraeder (2004), we decided to include trust as a variable in this study. Two types of trust are included: trust in the other and trust in the self. Trust in the other as assessor was included in the context of multi-rater i.e, multiple-perspective assessments, in which different others appraise an individual's performance. Trust in the self was examined and conceptualised as the trust people have in their own competence to assess a colleague.

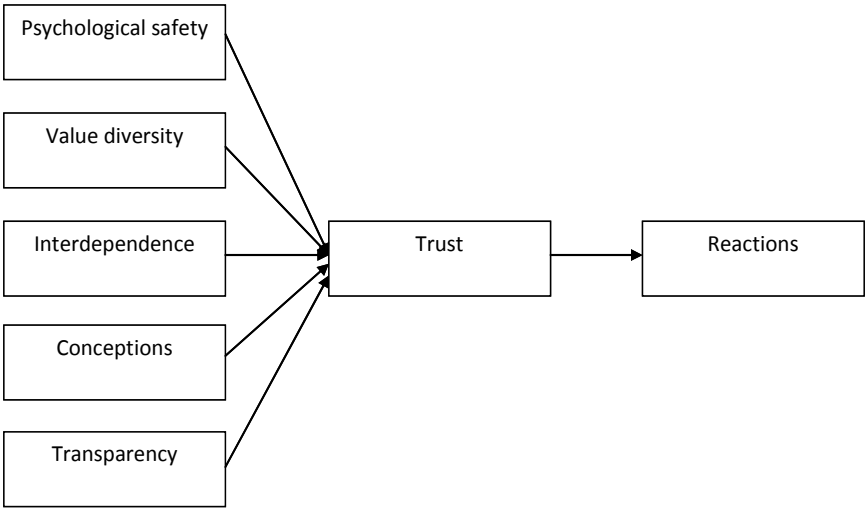


*Figure 2*
Explanatory model of this study

### 3.2 Interpersonal beliefs

The main purpose of implementing 360 degree feedback as a developmental tool is for co-workers to learn with and from each other. In short, 360 degree feedback is an instrument used mainly in collaborative and social settings within organisations. From the results of research on team learning in organisations we recognise three interpersonal beliefs that may be of importance for gauging the process of 360 degree feedback: perception of the psychological safety in the team, perceived value congruency within a team, and perceived interdependence between team members (Van den Bossche et al., 2006). Although we acknowledge that team learning settings differ from 360 degree feedback settings as they are used in organisations, the same conceptual underpinnings, i.e., regarding the social nature of learning, may be applied. Team learning is as a process of reflection and interaction in which team members actively acquire, process, and share knowledge and information in order to improve team performance (Rupert & Jehn, 2006). In a 360 degree feedback setting, information is shared using the processes of reflection and interaction among people who are related to each other in the work setting as well. Given these common feature of team learning and 360 degree feedback, and because of the evaluative character of the information, we can expect that interpersonal beliefs play a role in 360 degree feedback settings.

#### *3.2.1 Psychological safety*

Psychological safety can be described as the shared belief that a team or group of people is a safe environment for interpersonal risk-taking (Edmondson, 1999). The idea that psychological safety could influence the learning effects of 360 degree

feedback has been the result of the positive effect of psychological safety on learning and group effectiveness found in earlier studies (e.g., Edmondson, 1999; Van den Bossche et al., 2006). Psychological safety, for example, prevents teams from perceiving differences in viewpoints as disagreements; instead, psychological safety creates room for coming to grips with a problem and hence for collaborative learning. Team members' feelings of psychological safety affect their investment in learning from and with other team members. When 360 degree feedback is implemented as a developmental tool, psychological safety is a necessary condition for assessees if they are to invest in learning and professional development on the basis of the feedback results.

### *3.2.2 Value congruency*

Value congruency is defined as a similarity in opinion of what a team's task, goal, or mission should be (Jehn et al., 1999). In their study, Jehn et al. (1999) have shown that value congruency in teams should be high in order for teams to be effective. Integrating different perspectives and developing a shared understanding is crucial if teams are to perform well (Van den Bossche et al., 2006). Additionally, from feedback research we learn that goal setting is important for achieving positive reactions to feedback (e.g., Kluger & DeNisi, 1996). Moreover, educational research on the learning effects of peer assessment in higher education has shown that a shared understanding of learning goals within the peer group is necessary for learning effects to occur (e.g., Sivan, 2000). A shared understanding implies high value congruency, i.e., a large number of shared goals. In short, prior research indicates a potential relation between value congruency and reactions to 360 degree feedback, although evidence is scarce.

### *3.2.3 Interdependence*

Interdependence between members of a group may also affect responses and attitudes towards the group (e.g., Duimering & Robinson, 2009). Interdependence can be seen as "the division of labor within groups of departments" (Van der Vegt et al., 1998). Earlier research results have not been entirely clear about the relation between interdependence, performance, and feelings of responsibility in work group settings (Van der Vegt et al., 1998), but there is a consensus that it is of significant influence in work teams. Prior research on 360 degree feedback indicates that the source of the feedback affects assessees' reactions. These sources are peers and supervisors as well as clients. The more closely assessor and assessee have been working together, dependent on each other in the performance of tasks, the more the feedback source can be trusted or seen as valid.

### 3.3 Conceptions of assessment

During the past decade a number of studies have been conducted on conceptions of assessment, indicating the importance of these conceptions for the acceptance and validity of assessments. Thompson (1992, p. 130) considered conceptions "a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like." Furthermore, conceptions can be described as a framework through which one 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 learning (Hirschfeld & Brown, 2009); analyses of the influence of interpersonal beliefs on perceived learning from 360 degree assessments have shown that conceptions on assessment may filter the outcomes. It has been found that people generally perform better when their conceptions regarding a task are more positive (Brown et al., 2009), partly because they feel more competent at the task when they are more optimistic about it.

In the context of 360 degree feedback, a few studies refer to conceptions. A study by Atwater and Brett (2005), for example, has shown that individuals using feedback and viewing it in a more positive light showed more positive emotions afterwards than people who were less positive. We included employees' conceptions of 360 degree feedback in order to understand the role of these predictors on employees' reactions more fully.

### 3.4 Transparency

People can have trouble using assessments or feedback comments effectively because they find it difficult to interpret them correctly. Therefore, transparency of assessment procedures and criteria is important in order to improve people's awareness of the quality of their own performance (Sluijsmans et al., 2002), and might therefore influence the trust they have in their competence to assess someone else. Research has shown that in a peer-assessment setting good training in and explanation of this type of assessment does provide more transparency (Sluijsmans et al., 2004), but the relations between transparency and trust or transparency and reactions to feedback have not yet been investigated.

In the context of 360 degree feedback in the workplace, Van der Heijden and Nijhof (2004) refer to the relevance of transparency by arguing that an effective application of 360 degree feedback depends upon a careful formulation of criteria, and a thorough operationalisation of the concept to be measured. The importance of transparency is confirmed by McDowall and Fletcher (2004). They relate transparency to perceived fairness, which in turn is associated with trust and interpersonal beliefs. However, they argue that more research on the topic of fairness and

its interrelations with interpersonal beliefs is necessary (McDowall & Fletcher, 2004).

# 4 Method

## 4.1 Participants

Our sample consisted of 118 employees from four different organisations in The Netherlands. Three organisations were institutions from the non-profit sector: one psychiatric hospital (N = 20), one general hospital (N = 19) and one general health care institute (N = 48)). The fourth was a software engineering multinational (N = 31).

Following Bartram's (2004) suggestion, in order to cover a broad range of work organisations our study included employees from both profit and non-profit companies as well as companies that differ in size (from relatively small to multinational). In health organisations it is typically team or joint assessments that are used for improvement of performance, which therefore makes these a relevant setting for our study. In line with Atwater and Brett (2005) we believe that including participants from different organisations, thereby creating variety in task demands, will enhance the generalisability of our findings.

## 4.2 Procedure

All participants took part in 360 degree feedback procedures during the past year. More specifically, they were assessed by at least three out of four assessors in 360 degree feedback procedures (see Figure 1). All participants were acquainted with the process and had themselves also been assessors at least once in the past year. The questionnaire sets were distributed to the participants of this study by the researchers in order to prevent any confounding effects from supervisors or managers; the questionnaires had to be returned within one week. It consisted of eight scales, measuring: the *dependent variable* ´reactions to 360 degree feedback´; the *mediators* ´trust in the self´ and ´trust in the other´; and the *independent variables* ´psychological safety´, ´value congruency´, ´interdependence´, ´transparency´, and ´conceptions of 360 degree feedback´.

## 4.3 Instrumentation

The constructs identified in the conceptual framework of this study were measured by questionnaire scales taken from validated instruments. These scales have been developed and tested in several studies on peer assessment in higher education and team learning (see below). Measurements were done by 5-point Likert scales,

running from “1 = Totally Untrue” to “5 = Totally True”. A reliability analysis of instruments revealed coefficient alphas of more than .60 (see Table 2).

### 4.4 Trust

The questionnaire measuring the construct of trust was an adapted version of the ’assessment skill‘ scale by Sluijsmans et al. (2002); it measured trust in the self and in the other as an assessor. In the Sluijsmans et al. (2002) study the scale was used to measure trust in self-perceived assessment skills within a peer assessment setting. We adapted the scale for use in an in-company 360 degree feedback setting, and expanded it to measure perceived trust in the other as an assessor. Items included: ”I can judge whether my colleagues are doing their work well” (trust in the self as assessor) and ”My colleagues can judge whether others are doing their work well” (trust in the other as assessor). Both four-item scales proved to be reliable (Cronbach’s $\alpha$ = .70 (trust in the self); $\alpha$ = .83 (trust in the other)).

### 4.5 Interpersonal beliefs

The interpersonal beliefs involved were: psychological safety, value congruency, and interdependence.

#### *4.5.1 Psychological safety*

The scale measuring psychological safety was taken from Edmondson (1999) and consisted of seven items, for instance: ”People can raise difficult topics in this department.” Internal consistency is acceptable (Cronbach’s $\alpha$ = .68.).

#### *4.5.2 Value congruency*

The scale measuring value congruency was measured using six items (for example: ”All individual colleagues aim at shared goals”), with Cronbach’s $\alpha$ .66 for the scale. The items were adopted from a study by Jehn et al. (1999) which addressed the perceived similarities between team members on the team’s tasks, goals, and mission.

#### *4.5.3 Interdependence*

Items measuring interdependence were based on the scales developed by van der Vegt et al. (1998). The scale measuring task interdependence was adapted (e.g., ”I depend on my colleagues for information and advice”). Both scales, i.e., dependence of the self (with $\alpha$ = .80) and dependence of the other (with $\alpha$ = .78) consist of four items.

### 4.6 Conceptions of assessment

The scale measuring conceptions of 360 degree feedback, a shortened version of a questionnaire developed by Sluijsmans et al. (2002), consisted of ten items (e.g., ”360 degree feedback is informative”) and proved to be internally consistent (with $\alpha = .78$).

### 4.7 Transparency

The scale measuring transparency included five items and was developed by Van Gennip, Segers and Tillema (2006). Examples are: ”The goal of 360 degree feedback is clear to me” and ”The way in which we apply 360 degree feedback is clear to me.” Internal consistency was found to be acceptable (with $\alpha = .76$).

### 4.8 Dependent variable: Reactions to 360 degree feedback

The ‘reactions to 360 degree feedback’ scale was developed by Van Gennip et al. (2006), and conceptualised as ‘perceived improvement in performance’. This scale ($\alpha = .60$) contained three items, namely: ”It is easier to do my job because I assessed my colleagues”, “I am better at my job because we assessed each other”, and ”I learnt to be more critical of my own functioning through assessing the others.”

### 4.9 Data analysis

As a first step in the analysis of the data we conducted a descriptive and correlational analysis. Second, in order to test for direct effects on reactions to 360 degree feedback of interpersonal beliefs, conceptions, and transparency, stepwise regression analyses were used. In a first analysis, the impact of interpersonal beliefs on reactions to feedback was examined. A second analysis gauged the influence of interpersonal beliefs, together with conceptions and transparency, on reactions to 360 degree feedback. Third, we analysed the role of trust. In order to test for the mediating role of trust on reactions to 360 degree feedback a hierarchical regression analysis was used. This regression model consisted of three steps. In a first step we tested the effect of all independent variables on trust in the self and the other; the second and third steps contained all independent variables as predictors of reactions to 360 degree feedback, alternatingly excluding and including trust as a predictor variable.

# 5 Results

Means, standard deviations, and Pearson correlations of measured variables are presented in Table 1, together with scale reliabilities.

Table 1
*Means, standard deviations, alphas, and intercorrelation coefficients of variables*

| Variable | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Reactions | 3.6 | .69 | (.60) | | | | | | | | |
| 2. Conceptions | 3.9 | .48 | .35** | (.78) | | | | | | | |
| 3. Transparency | 4.1 | .56 | .27** | .49** | (.76) | | | | | | |
| 4. Trust in self | 3.8 | .50 | .30** | .48** | .57** | (.70) | | | | | |
| 5. Trust in other | 3.4 | .72 | .36** | .27** | .19* | .31** | (.83) | | | | |
| 6. Psychological safety | 3.8 | .58 | .20* | .18 | .15 | .15 | .37** | (.68) | | | |
| 7. Value congruency | 3.5 | .59 | .35** | .19* | .23* | .22* | .39** | .41** | (.66) | | |
| 8. Dependence self | 3.8 | .47 | .02 | .16 | .27** | .21* | -.08 | -.02 | .10 | (.80) | |
| 9. Dependence other | 3.6 | .67 | -.00 | .00 | -.00 | -.10 | .15 | -.02 | .00 | .30** | (.78) |

*p < .05; **p < .01; alphas are in parentheses

Some interesting findings can be noted here. First, the results from correlation analysis show that the participants' *reactions to feedback* are related to many of the independent variables included in this study: transparency of the assessment, conceptions of 360 degree feedback, trust, psychological safety, and value congruency. Only the participants' perceptions of dependence of the self and dependence of the other showed no correlations with reaction to 360 degree feedback. Second, *trust* in the self and the other as assessors is higher when the assessment process is perceived as more transparent. Moreover, trust in the self as assessor is significantly related to trust in the other as assessor. Third, the correlation analyses offer further insights into the *conceptions* participants have of 360 degree feedback. The conceptions of 360 degree feedback were positively related to perceptions of value congruency and transparency of the 360 degree feedback, and to the extent to which people trust themselves and the other as assessors. Fourth, of the *interpersonal beliefs* shared goals (i.e., high value congruency) and trust in the other as assessor were related to perceptions of psychological safety. In addition, high value congruency was significantly associated with trust in the other and the self, as well as with transparency. Finally, participants perceive the assessment as *transparent* when they trust themselves and the other as assessors (see Table 1).

### 5.1 The influence of interpersonal beliefs on employees' reactions to 360 degree feedback

We expected an influence of all interpersonal beliefs on reactions to 360 degree feedback. However, the two beliefs 'dependence on the self' and 'dependence on the other' do not show significant correlations. Results of stepwise regression analysis show that only value congruency is a significant predictor of reactions to 360 degree feedback ($\beta = .351$, $p < .001$). Employees with a high perceived value congruency (implying high agreement on shared goals) respond more positively towards 360 degree feedback.

### 5.2 The influence of conceptions and transparency on employees' reactions to 360 degree feedback

We expected some influence of conceptions and transparency on reactions to 360 degree feedback on the basis of previous research (e.g., McDowall & Fletcher, 2004; Atwater et al., 2000). Stepwise regression analysis revealed a significant portion of the variance of reactions to 360 degree feedback ($R^2 = .132$, $p < .001$) as explained by these variables. However, only 'conceptions of 360° feedback' was a significant predictor of reactions to 360 degree feedback ($\beta = .284$, $p < .05$).

### 5.3 Trust as a mediator variable

Considering the low to moderate correlations between transparency, conceptions, and interpersonal beliefs with regard to reactions to 360 degree feedback, we analysed the possible mediating effect of trust. A full model was tested with trust as a mediator, using regression analysis. First, all independent beliefs were entered in order to predict trust in the self and the other. Results indicate that transparency and conceptions predict degree of trust in the self as an assessor ($R^2 = .396$; $p < .01$). In contrast, trust in the other is predicted by the interpersonal beliefs ($R^2 = .287$; $p < .01$). Second, the independent variables were entered in order to predict reactions to 360 degree feedback excluding trust, and later including trust. Results show a partial mediation: the beta of value congruency decreases (but stays significant) when trust in the peer as assessor is included in the model. This implies the relation between value congruency and reactions to feedback is partially mediated by trust in the other as assessor.(see Table 2).

Table 2
*Regression analysis predicting trust*

| | Trust in self | Trust in other | Reaction to feedback (without trust) | Reaction to feedback (trust included) |
|---|---|---|---|---|
| Conceptions | .26*** | .18 | .25** | .20** |
| Transparency | .39*** | .07 | .10 | .05 |
| Psychological safety | .01 | .20** | .03 | -.01 |
| Value congruency | .07 | .28*** | .28*** | .22** |
| Dependence self | .10 | -.21** | -.09 | -.05 |
| Dependence other | -.12 | .21** | .03 | -.01 |
| Trust in self | | | | .08 |
| Trust in other | | | | .19* |
| Adjusted $R^2$ | .40 | .29 | .17 | .19 |
| F | 12.14*** | 7.43*** | 5.07*** | 4.52*** |

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

The results of our analyses revealed that more positive conceptions of 360 degree feedback implicate more positive reactions to 360 degree feedback. Further, trust in the other was found to affect reactions to 360 degree feedback. Moreover, an indirect effect of psychological safety, dependence of the self and dependence of the other is found, that is: the more employees perceive the interpersonal environment as safe and the more they feel able to depend on each other, the higher the degree of trust in the other as an assessor, and in turn, the more positive their reactions to 360 degree feedback.

## 6. Conclusion and discussion

Our study aimed at broadening our understanding of the interpersonal context in which peer assessment takes place within organisations where peer assessment is the heart of the 360 degree feedback system. More specifically, we analysed the separate and joint effects of trust and trust-related variables (interpersonal, transparency, and conceptions) on employees' reactions to 360 degree feedback on the basis of how they perceived improvement in functioning. More specifically, we studied the influence of trust, perceptions of interpersonal beliefs (psychological safety, value congruency, and interdependence), transparency of the feedback system, and conceptions of 360 degree feedback.

Our results seem partly to confirm previous research. First, although the role of interpersonal beliefs has been mainly examined in research on team learning, our study confirms that they are of significant importance in an assessment setting, which is inherently a social activity. Our study seems to indicate that value congruency predicts employees' reactions to feedback, and the perception of psychologi-

cal safety predicts the perception of trust. Second, our study may confirm the results of a study by Atwater and Brett (2005) showing that individuals with more positive conceptions of feedback showed more positive emotions following feedback than people with less positive conceptions. Third, the importance of assessment transparency, widely advocated in educational assessment research, was partly confirmed. Although the perception of transparency does not directly influence employees' reactions to feedback, it does seem to predict the trust they have in themselves as assessors. However, trust in the self as an assessor did not predict reactions to 360 degree feedback. Finally, we paid special attention to the role of trust in employees' reactions to 360 degree feedback. Although this variable is recognised as important, it has hardly been studied to this date. Our results indicate that trust in the other as assessor partially mediates the relation between value congruency and employees' reactions to feedback. The effects of trust in the other on reactions to 360 degree feedback seem to contradict the findings by Atwater and Brett (2005). However, their lack of results may be explained by not considering the role of interpersonal beliefs.

The study presented here should be informative for practitioners in the field of HRM and HRD who use 360 degree feedback as a tool to stimulate employees to invest in professional development. Given that positive reactions to such feedback are a necessary condition for success, organisations should encourage employees to develop trust in the other as assessor. This means that attention should be paid to establishing shared goals and creating a positive social climate where team members feel connected and safely dependent on each other. Moreover, making the 360 degree process as transparent as possible is a fruitful approach. Paying attention to sharing goals seems to be a strong tool, as it influences employees' reactions to feedback both directly and indirectly (i.e., by trust in the other as assessor).

Several implications for future research can be formulated. First, most of the participants in this study were employees in non-profit organisations. It would be interesting to measure whether factors behave differently in profit or non-profit organisations, using large samples in both settings. For example, a more competitive atmosphere in non-profit organisations may be expected to lead to different relations between people working together. This need not result in different predictive models. Differences between companies regarding the independent variables do not change the patterns of interaction between variables. Nevertheless, it would be interesting to repeat our study with large samples of employees in organisations that differ from each other on various parameters that might affect the variables distinguished. Large samples will make it possible to perform multi-level analyses in order to find clear evidence for the organisation effect.

Second, employees' reactions to feedback may influence how or whether they change their behavior in response to the feedback. It is a challenge for future re-

search to include performance measures operationalising employees' behavioral changes. To this date, research on the effects of interventions intended to support professional development is still struggling with the operationalisation and measurement of these effects. Some authors prefer a qualitative approach, using interviews to elicit whether and how employees change their professional behavior as a result of formal or informal learning opportunities in the workplace (e.g., Doornbos & Krak, 2001). Others use proxy measures such as employability to indicate effects of professional development interventions (e.g., Sanders & de Grip, 2004).

Third, the reliability of the three-item 'reactions to feedback' scale is moderate, so that the results of this study should be interpreted with caution. The development of a more extended scale may enhance its reliability. The benefit of using surveys as a research method is the possibility to question larger samples than in qualitative approaches, so that patterns in relations between variables can be explored. However, in a qualitative approach it is possible to acquire more in-depth information on the participants' feelings and reactions to interventional situations, and clarify unclear points by asking additional questions. The advantages of both methodologies may be brought together by combining both quantitative and qualitative approaches in future research.

This study aims to contribute to the discussion about interpersonal beliefs in the field of 360 degree feedback. We would like to point out that there may be differences between companies as to how employees perceive the independent variables. This need not result in different predictive models. However, regression analyses do indicate that interpersonal beliefs, transparency, conceptions and trust, directly or indirectly affect reactions of employees. This study indicates that interpersonal beliefs are of importance in a 360 degree feedback setting. We believe that such a multi-source feedback instrument will function best when the social environment is taken into account.

## References


Atwater, L. E., Waldman, D. A., Atwater, D., & Cartier, P. (2000). An upward feedback field experiment: Supervisors' cynicism, reactions, and commitment to subordinates. *Personnel Psychology*, *53*, 275-297.

Bailey, C., & Austin, M. (2006). 360 degree feedback and developmental outcomes: The role of feedback characteristics, self-efficacy and importance of feedback dimensions to focal managers' current role. *International Journal of Selection and Assessment*,14, 51-66.

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research project. *Assessment & Evaluation in Higher Education*, *27*, 427-441.

Becton, J. B., & Schraeder, M. (2004). Participant input into rater selection: Potential effects on the quality and acceptance of ratings in the context of 360-degree feedback. *Public Personnel Management, 33*, 23-32.

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.

Doornbos, A., & Krak, A. J. (2001). Learning processes and outcomes at the workplace; a qualitative research study. In: J. N. Streumer (Ed.), *Perspectives on learning at the workplace: Theoretical positions, organizational factors, learning processes and effects, Proceedings Second Conference HRD research and practice across Europe* (pp. 53-64).

Duimering, P. R., & Robinson, R. B. (2009). Effects of context on team behaviour. *International Journal of Human Resources Development and Management, 1*, 19-35.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*, 350-383.

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.

Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management, 32*, 325-351.

Hirschfeld, G. H. F., & Brown, G. T. L. (2009). Students' conceptions of assessment: Factorial and structural invariance of the SCoA across. *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.

Jellema, F., Visscher, A., & Scheerens, J. (2006). Measuring change in work behavior by means of multisource feedback. *International Journal of Training and Development, 10*, 121-139.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: Historical review, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.

Lievens, F., Harris, M., & Van Keer E. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology, 88*, 476-492.

London, M., & Smither, J. W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. *Personnel Psychology, 48*, 803-839.

Marton, F. (1981). Phenomenography – describing conceptions of the world around us. *Instructional Science*, *10,* 177-200.

Maurer, T. J., Mitchell, D. R. D., & Barbeite, F. G. (2002). Predictors of attitudes toward a 360-degree feedback system and involvement in post-feedback management development activity. *Journal of Occupational and Organizational Psychology*, *75*, 87-107.

McDowall, A., & Fletcher, C. (2004). Employee development: An organizational justice perspective. *Personnel Review, 33*, 8-29.

McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovation in Education and Training International*, *32*, 302-313.

Morgeson, F. P., Mumford, T. V., & Campion, M. A. (2005). Using research and practice to address 27 questions about 360-degree feedback programs. *Consulting Psychology Journal: Practice and Research, 57*, 196-209.

Rupert, J. and Jehn, K. A. (2006, September). *Team learning: The development and validation of a new typology.* Paper presented at the 10th International Workshop on Teamworking, Groningen, The Netherlands.

Sanders, J., & de Grip, A. (2004). Training, task flexibility and low-skilled workers' employability. *International Journal of Manpower, 25*, 73-89.

Sivan, A. (2000). The implementation of peer assessment: an action research approach. *Assessment in Evaluation, 2*, 193-213.

Sluijsmans, D. M. A., Brand-Gruwel, S., & Van Merrienboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment & Evaluation in Higher Education, 27*, 443-454.

Sluijsmans, D. M. A., Brand-Gruwel, S., Van Merriënboer, J. J. G., & Martens, R. (2004). Training teachers in peer-assessment skills: Effects on performance and perceptions. *Innovations in Education and Teaching International, 41*, 59-78.

Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology, 58*, 33-66.

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 Publishing Company.

Thornow, W. W. (1993). Perceptions or reality: Is multi-perspective measurement a means or an end? *Human Resource Management, 32*, 221-229.

Tillema, H. (2001). *Assessment van competenties: van beoordelen naar ontwikkelen*, Mechelen: Kluwer.

Tynjälä, P. (2006). *Perspectives into Learning at the Workspace.* Paper presented at the chair for educational innovation and cooperation, University of Antwerp, Belgium.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 66,* 249-276.

Van den Bossche, P., Gijselaers, W., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments. Team learning beliefs & behaviors. *Small Group Research, 37*, 490-521.

Van der Heijden, B. I. J. M., & Nijhof, A. H. J. (2004). The value of subjectivity; problems and prospects for 360-degree appraisal systems. *International Journal of Human Resource Management*, *15*, 493-511.

Van der Vegt, G., Emans, B., & Van de Vliert, E. (1998). Motivating effects of task and outcome interdependence in work teams. *Group & Organization Management, 23*, 124-143.

# CHAPTER 6
# Summary, general conclusion and discussion

During the past two decades, there has been a lot of debate on the powerful role of assessment to support student learning. Former research has indicated that for assessment to support student learning, various conditions have to be met (e.g., Assessment Reform Group, 2002; Black & Wiliam, 1998; James & Pedder, 2006). These conditions have in common the explicit attention for the active role of students. When students are involved in the various steps of the assessment cycle, significant increases in learning gains are reached. In this respect, the value of peer assessment has been argued. However, years after its implementation in educational practices, there still remain a lot of unanswered questions with respect to the nature of peer assessment as well as the processes that support strong peer assessment environments. With respect to the former, descriptions in literature of peer assessment arrangements show that there are large differences in the structural features of peer assessment. Peer assessment arrangements vary in the organization of the assessment, the interactions within the peer assessment, and the composition of the feedback group (Van Gennip, Segers, & Tillema, 2009). As an example of differences in interactions, peer assessment can differ as to level of privacy (anonymous, confidential, public) and contact between assessor and assessee (from a distance or face to face). With respect to the latter - the processes that support strong peer assessment environments - the inconclusive results of peer assessment effect studies indicate that, in order for peer assessment to enhance students' learning gains, attention has to be paid to the social and interpersonal aspects of peer assessment, inasmuch as they can influence under what conditions students accept each other as assessors of their learning, and trust the appraisal outcomes of peer assessment.

The research presented here focused on how the social context in which peer assessment takes place is related to the learning effects of peer assessment. We

wanted to identify interpersonal beliefs in peer assessment settings that influence the peer assessment process on the one hand, but could also be seen as predictors of the effect of peer assessment interventions. In order to operationalise the interpersonal context of peer assessment, and herein inspired by research on team learning, we addressed different beliefs that have been proved to influence the extent to which team members learn from and with each other: psychological safety, trust, value congruency, and interdependence.

This dissertation presents four studies that question parts of the interpersonal context of peer assessment: one systematic literature study on interpersonal beliefs in and structural features of peer assessment and three empirical studies. In the systematic literature study of this dissertation (chapter 2), we presented a structural model of analysis indicating different interpersonal beliefs, as well as structural features of peer assessment that might affect outcome measures of peer assessment (achievement, perceptions of learning benefits and conceptions of assessment). This review is followed by three experimental studies. Study 1 (chapter 3) is a first exploration of the effects of interpersonal beliefs and focuses on the differences in interpersonal beliefs between two peer assessment settings and a teacher assessment setting. Furthermore, the use and effect of an intervention on interpersonal beliefs is studied. Study 2 (chapter 4) examines interpersonal beliefs more in depth and studies which interpersonal beliefs lead to perceived learning from peer assessment. Finally, study 3 (chapter 5) adds a new peer assessment setting to this dissertation, and attempted to find out whether the same interpersonal processes play a role in a peer assessment setting implemented in a professional environment as do in an educational one.

## 1 An overview of the results

### 1.1 A review on the literature addressing peer assessment

The literature review (chapter 2) aimed to present an overview of empirical studies evidencing the effect of peer assessment on student learning and the role of the social context of peer assessment to realize this effect. With respect to the latter, in line with a large body of research on the social context of team learning, we focused on the structural features of the peer assessment arrangements as well as the interpersonal beliefs of the students involved in the peer assessment. The systematic literature search resulted in 15 studies conducted since 1990 dealing with effects (performance or perceived learning gains) of peer assessment. Our analysis reveals that, although peer assessment is inherently a social process, only four out of fifteen studies addressed interpersonal beliefs, more precisely, psychological

safety and trust. However, no evidence was presented on their role in enhancing learning gains.

Further, comparing the studies with respect to structural features of peer assessment arrangement reveals that, although the differences between the studies are significant, there seems to be no relation with the occurrence (or non-occurrence) of learning benefits. Moreover, there is a lack of research explicitly relating features of the peer assessment setting to learning gains.

The results of this review seem to indicate that research on peer assessment from an interpersonal perspective is still in its infancy and deserves more attention, which strengthened us in our conviction of the urgency of our empirical studies. Further, the structural features of peer assessment formats are already a recognized subject of research, but these have not been related to the learning effects of peer assessment. Therefore, research designs should be clear and well-grounded if any conclusions are to be drawn on the basis of these structural features.

### 1.2 The role of interpersonal beliefs in peer assessment

Both empirical study 1 and 2 (chapters 3 and 4) focused on the role of interpersonal beliefs in peer assessment: we wanted to examine our hypothesis that interpersonal beliefs change in a peer assessment setting. Team learning research already evidenced that interpersonal beliefs are a precondition for group learning outcomes (e.g., Edmondson, 1999; Van den Bossche, Gijselaers, Segers, & Kirschner, 2006). Given an increase in learning gains is a primary goal of peer assessment as well, studying the role of students' interpersonal beliefs is highly relevant.

The first question in the exploration of the role of interpersonal beliefs was whether participation in a peer assessment intervention results in a change in perceptions of interpersonal beliefs (psychological safety, value congruency) during the peer assessment practice. In order to find out whether this increase in interpersonal beliefs is indeed an effect of the peer assessment intervention itself, we compared two peer assessment settings with a teacher assessment setting (as a baseline condition). Results of study 1 (chapter 3) show that beliefs of value congruency and psychological safety are higher at the end of the project in a peer assessment condition than in the teacher assessment condition.

Given the indication of the role of interpersonal beliefs in peer assessment in study 1 (chapter 3), based on findings of team learning research, two variables were added to study 2 (chapter 4): trust and interdependence. Comparing peer assessment students' scores on the interpersonal beliefs at the end of the project with those from the teacher assessment students revealed, comparable to study 1, that psychological safety and value congruency were higher in the experimental group: students in a peer assessment setting feel significantly safer and perceive more unanimity in goals than students in a teacher assessment setting. Trust and

interdependence, however, were not perceived differently by students from the experimental group compared to the teacher-assessed students.

Recognizing the result that beliefs of psychological safety and value congruency are higher in a peer assessment setting compared to a teacher assessment setting, we wanted to check whether an intervention aimed at stimulating these interpersonal beliefs would lead to an even higher increase in those beliefs. Therefore, we compared a peer assessment$^{+}$ arrangement (including an intervention) with a regular peer assessment arrangement. Results showed that there are no differences between peer assessment and peer assessment$^{+}$ condition for psychological safety as well as value congruency. Further, results show that differences in perception of value congruency in the peer assessment$^{+}$ condition appear before the reflection session has taken place. Given this result as well as the observed differences in interpersonal beliefs between the peer assessment conditions and the teacher-based condition, it seems that implementing peer assessment in itself leads to more positive interpersonal beliefs (psychological safety and value congruency). For value congruency, involving students in the first stage of the peer assessment process (goals and purposes formulation) seems to be beneficial and, adding a reflection session after this first stage, seems to have no value added.

Finally, as a next step in understanding the role of interpersonal beliefs in peer assessment, we intended to explore how interpersonal beliefs are related to students' conceptions of peer assessment and to students' learning gains (as perceived by the students themselves and expressed by their performance). Results of study 2 (chapter 4) showed that some interpersonal beliefs indeed play a significant role in peer assessment settings and are of influence when we consider the perceived learning of students: Students' perception of learning effects are directly influenced by two interpersonal beliefs, namely the belief of value congruency and conceptions of peer assessment, which play an important role. Conceptions students hold who have no prior experiences with peer assessment predict the learning effects they experience. Moreover, their conceptions of assessment are influenced by the belief of psychological safety and trust in themselves and the other as well as by their belief of value congruency. Moreover, the results indicated a full mediation effect of conceptions of peer assessment regarding trust in the self as an assessor.

### 1.3 Changing the setting: Reactions to 360 degree feedback

In the last empirical study of this dissertation, we chose to examine the role of interpersonal beliefs in a peer assessment setting where peer assessors have prior experiences with peer assessment and where feedback is the core purpose. Peer assessment was part of the 360 degree feedback system. The aim of this study is to provide insight into the interpersonal beliefs that contribute to the employees'

reactions to 360 degree feedback in terms of learning effects: trust, psychological safety, value congruency, and interdependence. Furthermore, conceptions of assessment were included as predictor, given the participants in this study have prior experiences with peer assessment as part of the 360 degree feedback system they regularly participate in. Moreover, given prior research on 360 degree feedback systems stress the importance of transparency, this variable was also taken into account as predictor.

Results reveal that trust in the other as assessor partly mediates the relation between value congruency and reactions to 360 degree feedback. Conceptions and transparency are predictors of trust in the self, which does not predict reactions to 360 degree feedback. Finally, the interpersonal beliefs of psychological safety and interdependence have an indirect effect on reactions to 360 degree feedback: they are predictors of trust in the other, which in turn predicts reactions to 360 degree feedback .

An important finding in this study is that trust is clearly a divided variable: trust in the self and trust in the other are predicted by separate and different independent beliefs. Conceptions and transparency predict trust in the self, while the interpersonal beliefs of psychological safety, value congruency, and interdependence in turn are predictors of trust in the other.

## 2 Implications for practice

### 2.1 Implementing peer assessment as a tool for learning

The present research provides some insights that have implications for practice. Since the dissertation is based on samples in educational as well as professional practice, we will recall implications for teachers in education as well as managers in the workplace. All three empirical studies examined the role of interpersonal beliefs in a peer assessment setting. The results of the empirical studies deliver content for the empirical discussion of peer assessment: peer assessment as such seems to stimulate the beliefs in some interpersonal variables.

Our findings suggest that peer assessment is a powerful learning environment if certain conditions are taken into account. The classroom peer assessment setting used in the studies in this dissertation, involves students from the first steps in the assessment cycle. This means they are involved in collaboratively formulating the goals and the criteria of the peer assessment instead of only involving them at the rating stage of the assessment cycle. In the setting of our studies, the results indicate in classroom settings were students have no prior experiences with peer assessment, implementing a peer assessment arrangement where students are involved from the first stage of the assessment cycle, influences their beliefs of psy-

chological safety as well as value congruency. In addition, both beliefs influence the conceptions students hold at the end of a peer assessment intervention. In turn, these conceptions are significant predictors of the extent to which students experience learning effects by participating in peer assessment.

The latter is important as, when students do not experience peer assessment as a learning tool, the probability that they accept the feedback given and act upon, is high. This in turn will negatively influences actual learning gains.

For organizations where employees have already experiences with peer assessment , our finding suggest that given the relation between trust and reactions to 360 degree feedback, organizations should encourage employees to develop trust in the other as assessor. However, recognizing that interventions are not needed to strengthen interpersonal beliefs, the key to success of 360 degree feedback is to involve employees in the different steps of the assessment cycle. Moreover, making the 360 degree process as transparent as possible is a fruitful approach: the results of study 3 indicate the importance of transparency of the process of assessment as well. This can be achieved more easily when involving employees in the process of designing and carrying out the assessment.

### 2.2 Interpersonal beliefs and the quality of peer assessment

The confirmation of the role of interpersonal beliefs in peer assessment settings leads to the issue of *quality* of peer assessment. As we stated at the start of our dissertation - and this is confirmed by the results of our studies - peer assessment is a social activity, which implies an important role for interpersonal beliefs. Given this role, what does this say about the quality of assessment? The rethinking of the nature of assessment and the shift to assessment for learning made the use of the traditional criteria of validity and reliability no longer suitable for appraisal of quality of assessment (e.g., Linn, Baker, & Dunbar, 1991). We suggest that when teachers want to preserve quality of peer assessment as a tool for learning, interpersonal beliefs as studied in this dissertation should be recognized and considered as an aspect of quality. This is in accordance with earlier research of Tillema et al. (2010), who formulated three indicators to gauge quality assurance in assessment tasks, based on previous reviews. These indicators are authenticity (‘what’ criteria), transparency (‘how’ criteria) and generalisability (warranty criteria). They examined how often these criteria were taken into account in the different steps of the construction, delivery and decision-making of assessment tasks. When we relate these criteria to the interpersonal beliefs that we describe in this dissertation, authenticity and transparency especially are strongly related to the social environment of peer assessment: authenticity relates to student involvement and self-directedness, while transparency includes among other things the concept of fairness, which in turn is linked to psychological safety.

The results of this study show that most quality criteria were taken into account in the step of scoring (Tillema et al., 2010). Additionally, they observed that in the construction and administration of assessment for learning, transparency (i.e. fairness) and meaningfulness are important factors. These factors are closely related to the interpersonal environment of the specific situation of peer assessment: individual transparency for example does not automatically imply value congruency among students. This means that quality of peer assessment is not only determined by the involvement of individual students, but by adapting between students as well. Tillema et al. (2010) suggest a further development of quality criteria towards a more robust framework; we would like to add the involvement of interpersonal beliefs to this suggestion as a necessary addition to the discussion on quality criteria in assessment for learning.

## 3 Limitations and next steps in research

The results of the studies presented in this dissertation contribute to the understanding of the role of interpersonal beliefs in peer assessment. However, we see this dissertation as a starting point for research on the social context of peer assessment, and here take the liberty to suggest study ideas for future research.

The studies in this dissertation were carried out in a specific vocational educational setting. All students were male and worked together in project teams in which they gave each other feedback. Students in the studies had little to no experience with peer assessment. Given the growing implementation of peer assessment in educational and professional settings, and the fact that there are some indications that prior experience with peer assessment has a positive impact on its effect, we have to ask the question whether prior experience with peer assessment influences the role of interpersonal beliefs. This includes prior training, making the students confident with the objectives of peer assessment, general organization and procedures, developing and using criteria and the process of giving and receiving feedback to and from peers. Topping (1998) as well as Falchikov and Goldfinch (2000) conclude that a systematic investigation of the effects of repeated experience of peer assessment is an important topic for future research.

However, the positive effects of prior experience are all formulated in a conditional way: more experience with or knowledge of certain processes in peer assessment make peer assessment more successful. In the light of this dissertation, we do not expect prior experience to influence interpersonal beliefs as such, and the results of our studies in particular. Because every assessment situation is different, with different peers, the interpersonal environment will be unknown as well. The process of creating psychological safety and value congruency has to be repeated in every new peer assessment situation. Therefore, we expect interper-

sonal beliefs to play a significant role in every new peer assessment setting. However, prior experiences might help run these processes more smoothly: when students are more confident with the objectives of peer assessment, general organisation and procedures of peer assessment and developing and using criteria because they have prior experience, psychological safety and value congruency might be established sooner. This would make a nice object of future research, for example longitudinal research with individual as well as group or team measures.

Further suggestions for future research include, given the fact that results from our studies are all based on self-reports of students, a repetition of our research with data from teachers, for example. Second, in order to be able to explore our quantitative results more in depth, further, more qualitative research is needed. We would be especially interested to know *in which way* interpersonal beliefs influence the way students use and accept peer assessment, therefore influence learning gains. As stated before, more sensitive performance measures are necessary to fulfil that last goal. Third, the qualitative research can be supported by quantitative research examining the mediator factors which are responsible for learning gains more extensively. Research on team effectiveness, where most interpersonal beliefs of this dissertation are adopted from, for example examine the role of knowledge sharing and co-constructural knowledge in this respect (e.g., van den Bossche, 2006). Nevertheless, this dissertation embodies the idea that peer assessment is indeed an interpersonal process, which is a direction future research should explore more fully.

## References


Assessment Reform Group (2002). *Assessment for learning: 10 principles*. Cambridge, University of Cambridge School of Education.

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*, 7-74.

Butler, S. A., & Hodge, S. R. (2001). Enhancing student trust through peer assessment in physical education. *Physical Educator 58*, 30-42.

Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*, 350-381.

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.

James, M., & Pedder, D. (2006). Professional learning as a condition for assessment for learning. In J. Gardner (Ed.), *Assessment for learning: theory, policy and practice* (pp.27-43). London, England: Sage.

Keaten, J. A., & Richardson, M. E. (1992). *A field investigation of peer assessment as part of the student group grading process.* Paper presented at the Western Speech Communication Association Convention, Albuquerque, NM.

Linn, R., Baker, E., & Dunbar, S. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20,* 16-21.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academic Press.

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

Tillema, H., Leenknecht, M., & Segers, M. (2010). Assessing assessment quality: criteria for quality assurance in design of (peer) assessment for learning - a review of research studies. *Studies in Educational Evaluation, 37*, 25-34.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *66*, 249-276.

Van den Bossche, P., Gijselaers, 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 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.

# About the author

## 1 Biography

Nanine van Gennip (1979) graduated in Educational Psychology at the Faculty of Psychology of the Maastricht University. In 2004 she started her career as a Ph.D. student at the Department of Educational Studies at the Leiden University, where she worked on her thesis on 'Assessing together: Peer assessment from an interpersonal perspective'. Besides her work as a Ph.D.-student, she also was a part-time lecturer at this department.

From 2008-2010, in addition to her work at the Leiden University, Nanine van Gennip worked as a consultant for educational institutions at consultancy firm Organise to Learn. Since 2011 she has been working as a senior consultant in an institution for vocational education (Rijn IJssel).

## 2 List of publications and presentations

### Articles in refereed journals

Van Gennip, N., Gijbels, D., Segers, M., & Tillema, H. (2010). Reactions to 360° feedback: the role of trust and trust-related variables. International Journal of Human Resources Development and Management, 10, 362-379.

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.

Van Gennip, N. (2009). Invoering functiemix: Draagvlak noodzakelijk dus meer tijd geboden. *School Zakelijk, 3*, 13-14.

Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a team learning perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4,* 41-54.

Van Gennip, N., Tillema, H., & Segers, M. (2008). Peer assessment in teams. *Opleiding & Ontwikkeling, 9*, 19-22.

Van Gennip, N. (2008). Succesvolle implementatie van peer assessment. *School Zakelijk, 5*, 12-13.

### Chapter in book

Van Gennip, N., & Van der Wal-Maris, S. (2011). De inzet van peer assessment. In: J. Castelijns, M. Segers & K. Struyven (Eds.), *Evalueren om te leren* (pp. 101-119). Bussum, The Netherlands, Coutinho.

## Papers


Van Gennip, N., Reamdonck, I., Segers, M., & Tillema, H. (2009, August). *360 degree feedback as a tool for learning: The role of self directed learning orientation and interpersonal variables*. Paper presented at the 12th Biennial EARLI Conference, Amsterdam, The Netherlands.

Van Gennip, N., Gijbels, D., Segers, M., & Tillema, H. (2008, March). *Unraveling the interpersonal variables that contribute to the perceived learning effect of 360 degree feedback*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), New York, USA.

Van Gennip, N., Segers, M., & Tillema, H. (2007, August). *The role of psychological safety, value diversity, clarity, dependence and trust in a peer assessment setting*. Paper presented at the 12th Biennial EARLI Conference, Budapest, Hungary.

Sluijsmans, D., Van Gennip, N., Gielen, S., Phielix, C., Prins, F., Strijbos, J. W., Van de Watering, G., & Van Zundert, M. (2007, November). *Guidelines for integrating peer assessment in team work: Assuring psychological safety, group awareness and individual responsibility*. Workshop performed at the second European Practice Based and Practitioner Research (PBPR) Conference on Learning and Instruction, Maastricht, the Netherlands.

Van Gennip, N., Segers, M., Tillema, H., & Gijbels, D. (2006, August). *Effectiveness of peer assessment in education and professional organizations*. Paper presented at the Third Biennial Joint Northumbria/EARLI SIG Assessment Conference, Northumbria, UK.

Van Gennip, N., Segers, M., Tillema, H., & Gijbels, D. ( 2006, June). *The effectiveness of 360° feedback*. Paper presented at the 13th Educational Innovation in Business Education (EDiNEB) international conference, Lisbon, Portugal.

Van Gennip, N., Segers, M., & Tillema, H. (2006, Mei). *Peer assessment in het Middelbaar Beroepsonderwijs*. Paper gepresenteerd op de Onderwijs Research Dagen, Amsterdam, Nederland.

Van Gennip, N., Segers, M., Tillema, H., & Gijbels, D. (2006, Mei). *De effectiviteit van 360 graden beoordeling*. Paper gepresenteerd op de Onderwijs Research Dagen, Amsterdam, Nederland.

Van den Bossche, P., Van Gennip, N., Gijselaers, W., & Segers, M. (2005, August). *Collaborative learning as constructive conflict: Building shared mental models*. Paper presented at the 11th Biennial EARLI Conference, Nicosia, Cyprus.

Van Gennip, A. E., Segers, M. S. R., & Tillema, H. H. (2005, August). *Peer assessment as a tool for learning: a meta-analysis*. Paper presented at the 11th Biennial EARLI Conference, Nicosia, Cyprus.

Van den Bossche, P., Van Gennip, N., Gijselaers, W., & Segers, M. (2005, June 8-10). *Harvesting diversity: Building shared mental models by dealing constructively with conflict (poster).* Paper presented at the KNAW/EASP symposium Social Identity in Organizations, Amsterdam, The Netherlands.

Van den Bossche, P., Van Gennip, N., Gijselaers, W., & Segers, M. (2005, April). *Effects of diversity on conflict and group performance in professional work groups.* Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Montreal, Canada.

Van Gennip, N., Van den Bossche, P., Gijselaers, W., & Segers, M. (2004, June). *Team learning: Dealing with diversity*. Paper presented at the 11th Educational Innovation in Business Education (EDiNEB) international conference, Maastricht, The Netherlands.

# Samenvatting / Dutch summary

De verantwoordelijkheid die mensen al dan niet nemen voor hun eigen leven is een centraal thema op dit moment, zowel binnen de politieke als de publieke discussie: we gaan steeds meer toe naar een samenleving die vertrouwt op de eigen verantwoordelijkheid. Dit heeft tevens geleid tot een herformulering van het begrip leren: leren betekent dat een lerende de eigen kennis construeert, welke hij of zij kan gebruiken als een mechanisme om de wereld te interpreteren en complexe problemen mee op te lossen. Lerenden moeten zelfstandig zijn, en voortdurend bereid om hun kennis te gebruiken en verbreden. Ook hebben ze metacognitieve vaardigheden nodig om te kunnen reflecteren op hun eigen perspectieven, en die van anderen. Het is nodig effectieve strategieën te ontwikkelen om het eigen leren te plannen en te monitoren (Segers, Dochy, & Cascallar, 2003).

De veranderingen in het denken over leren hebben ook veranderingen teweeg gebracht in het denken over beoordelen. Zowel binnen als buiten de onderwijswetenschappen is men het er over eens dat het beoordelingsproces in lijn zou moeten zijn met het leer- en instructieproces (Birenbaum & Dochy, 1996). Dit heeft geleid tot de zogeheten evaluatiecultuur. Nauw verbonden met deze cultuur is het perspectief dat er een verandering dient op te treden van 'Evalueren van het leren' naar 'Evalueren om te leren'. Evalueren om te leren is als volgt geformuleerd door de Assessment Reform Group: "Evalueren om te leren is het proces van zoeken naar en interpreteren van bewijzen van het leerproces en de leerproducten, op basis waarvan de lerende en zijn/haar leerkracht kan besluiten waar de lerende is in zijn/haar leerproces, waar de lerende heen moet, en hoe deze dit het beste kan bereiken" (Assessment Reform Group, 2002, pp. 1-2).

Om optimaal te profiteren van nieuwe leeromgevingen is het noodzakelijk meer geïntegreerde beoordelingsmethoden te hanteren. De belangrijkste kenmerken van de evaluatiecultuur zijn dat evaluaties zijn gebaseerd op meerdere evaluatiemethoden en dat evaluatie een geïntegreerd onderdeel is van het onderwijs- en leerproces. Daarenboven wordt het belang beargumenteerd van de actieve rol van studenten, waarbij betrokkenheid in het evaluatieproces de kans vergroot dat de evaluatie zal leiden tot meer en diepgaandere kennis en inzichten bij de studenten. Actieve betrokkenheid impliceert dat niet het evaluatieproces wordt gezien als een sociaal, interactief gegeven. De interactie tussen leerlingen onderling en tussen leerlingen en leerkrachten in het evaluatieproces wordt gezien als waardevol voor het realiseren van leerwinst.

Het is precies deze argumentatie die geleid heeft tot het uitgangspunt van dit proefschrift: aangezien het opnemen van een actieve rol in de evaluatie door studenten bijdraagt tot leerwinst, is aandacht voor de interpersoonlijke context van evaluatie noodzakelijk. Dit is zeker het geval in peer evaluatiesettings, waarin studenten elkaar evalueren met als doel het leren te bevorderen.

Geïnspireerd door onderzoek bij teamleren hebben we de rol van verschillende interpersoonlijke opvattingen onderzocht, welke eerder hebben laten zien invloed te hebben op de manier waarop teamleden met en van elkaar leren: psychologische veiligheid, unanimiteit in doelen, onderlinge afhankelijkheid en vertrouwen (Edmondson, 1999; Lingard, Reznick, Espin, Regehr, & DeVito, 2002; Van den Bossche, Gijselaers, Segers, & Kirschner, 2006).

## 1 Peer evaluatie als middel om leren te ondersteunen

Verschillende manieren van evaluatie die Evalueren om te leren ondersteunen zijn de afgelopen tijd geïmplementeerd. Daarbij heeft ook peer evaluatie groeiende aandacht gekregen, zowel in het wetenschappelijk onderzoek, als in de dagelijkse onderwijspraktijk. De context van het onderzoek binnen dit proefschrift is *peer evaluatie* als een middel om leren te ondersteunen.

Bij peer evaluatie is het interpersoonlijke aspect van leren als het ware natuurlijk ingebed. Het onderliggende mechanisme dat zorg draagt voor het leren van peer evaluatie werkt als volgt: doordat studenten een actieve rol hebben in het analyseren en evalueren van hun eigen leerproces, hebben ze een hogere motivatie en bereidheid om verder te ontwikkelen en groeien. Daarnaast betekent peer evaluatie automatisch interactie en dialoog tussen lerenden, waarbij het geven van feedback de belangrijkste activiteit is. Om die reden zien wij peer evaluatie als een interpersoonlijk leerproces. Het definiëren van peer evaluatie als een interpersoonlijk proces betekent dat we rekening moeten houden met de interpersoonlijke context wanneer we de leereffecten van peer evaluatie willen begrijpen. Gezien de grote verschillen in implementatie van peer evaluatie op het gebied van de organisatie van de evaluatie, de interacties binnen peer evaluatie en de samenstelling van de feedback groep (Van Gennip, Segers, & Tillema, 2009) kijken we naast de interpersoonlijke benadering ook naar de kenmerken van de peer evaluatieomgeving.

Samenvattend onderzoeken we in dit proefschrift in hoeverre peer evaluatieomgevingen de interpersoonlijke opvattingen van lerenden beïnvloeden (psychologische veiligheid, unanimiteit in doelen, onderlinge afhankelijkheid en vertrouwen) en als resultaat daarvan leidt tot een hogere waargenomen leeropbrengst (schoolomgeving) of prestatie (werkomgeving).

## 2 Overzicht van de studies

Dit proefschrift presenteert vier studies die allen een deel van de interpersoonlijke context van peer evaluatie onderzoeken: een systematisch literatuuronderzoek en drie empirische studies. De eerste empirische studie is een verkenning van interpersoonlijke opvattingen in een peer evaluatieomgeving, en vergelijkt een peer evaluatieomgeving met een traditionele docentgestuurde evaluatieomgeving De tweede empirische studie gaat een stap verder, en onderzoekt welke interpersoonlijke opvattingen invloed hebben op waargenomen leeruitkomsten als gevolg van peer evaluatie. De derde en laatste empirische studie is tenslotte uitgevoerd in een andersoortige peer evaluatieomgeving, namelijk 360 graden feedback, en onderzoekt of in deze werkomgeving dezelfde interpersoonlijke opvattingen een rol spelen als in de schoolomgeving.

### 2.1 Literatuuronderzoek

Het doel van de literatuurstudie is het in beeld krijgen van een overzicht van het effect van peer evaluatie op het leren van studenten, aangetoond in eerdere empirische studies, en de rol van de interpersoonlijke context daarbij. Hierbij ligt de focus op zowel de kenmerken van de peer evaluatieomgeving (zoals plaats en tijd van de evaluatie en samenstelling van de evaluatiegroep), als op de interpersoonlijke opvattingen. De resultaten van deze studie laten zien dat slechts 15 empirische studies zijn uitgevoerd sinds 1990 die de effecten van peer evaluatie trachten te meten. Slechts vier van deze studies kijken ook naar de interpersoonlijke context van de peer evaluatie. We vinden echter geen bewijs voor de rol van deze variabele in de leeropbrengsten. Tenslotte blijkt er uit de overzichtsstudie dat er geen verband lijkt te zijn tussen de structurele kenmerken van de peer evaluatieomgeving en de leeropbrengsten. Echter, hier geldt dat het aantal onderzoeksstudies dat de kenmerken van de peer evaluatieomgeving koppelt aan het verwerven van leeropbrengsten erg beperkt is en veralgemeniserende uitspraken niet toelaten.

Deze literatuurstudie laat zien dat het onderzoek naar interpersoonlijke opvattingen in peer evaluatieomgevingen nog in de kinderschoenen staat. Dit heeft ons gesterkt in onze overtuiging van de noodzaak van onze empirische studies. De kenmerken van de peer evaluatieomgeving zijn wel onderwerp van onderzoek geweest, maar deze kenmerken zijn nog nauwelijks verbonden aan de leeropbrengsten.

### 2.2 De rol van interpersoonlijke opvattingen bij peer evaluatie

De eerste en de tweede empirische studie (hoofdstukken 3 en 4) onderzoeken de rol van interpersoonlijke opvattingen van studenten in een peer evaluatie setting. De eerste vraag hierbij is of er een verandering optreedt in de interpersoonlijke

opvattingen wanneer studenten participeren in een peer evaluatieomgeving. De resultaten van studie 2 (hoofdstuk 4) laten inderdaad een verandering zien in de perceptie van de interpersoonlijke opvattingen: studenten in een peer evaluatie groep ervaren aan het einde van een project meer unanimiteit in doelen en meer vertrouwen in de peer als een assessor dan aan het begin van het project. Echter, scores op psychologische veiligheid, onderlinge afhankelijkheid en vertrouwen in zichzelf als assessor verschillen niet.

Om te onderzoeken of dit verschil in perceptie ook komt door de peer evaluatieomgeving hebben we vervolgens een peer evaluatieomgeving vergeleken met een docentgestuurde evaluatieomgeving. Zowel studie 2 als studie 3 laten zien dat de opvattingen van studenten over psychologische veiligheid en unanimiteit in doelen positiever zijn in de peer evaluatiegroep dan in de docentgestuurde evaluatiegroep: studenten in een peer evaluatiegroep voelen zich veiliger en percipiëren meer unanimiteit in doelen dan studenten die door hun docent zijn beoordeeld. Studie 3 laat zien dat er geen verschillen optreden ten aanzien van onderlinge afhankelijkheid en vertrouwen.

Gezien het resultaat dat psychologische veiligheid en unanimiteit in doelen hoger is in een peer evaluatieomgeving dan in een docentgestuurde evaluatieomgeving, willen we weten of een interventie, gericht op het stimuleren van deze interpersoonlijke opvattingen ook leidt tot een hogere waarde ervan. Om die reden vergelijken we een peer evaluatie$^{+}$ conditie (met interventie) met een 'gewone' peer evaluatie conditie. De resultaten laten zien dat studenten inderdaad meer psychologische veiligheid en unanimiteit in doelen ervaren in de peer evaluatie$^{+}$ conditie. Echter, dit verschil verschijnt al voordat de interventie heeft plaatsgevonden. De resultaten lijken erop te duiden dat de peer evaluatie interventie zoals wij die hebben geïmplementeerd op zichzelf al krachtig genoeg is om de interpersoonlijke opvattingen te doen toenemen: het werpt zijn vruchten af om studenten te betrekken in de verschillende stappen van het peer evaluatieproces.

Tenslotte hebben we onderzocht hoe de interpersoonlijke opvattingen van studenten verband houden met hun concepties over peer evaluatie en leeropbrengsten. De resultaten van studie twee (hoofdstuk 4) laten zien dat interpersoonlijke opvattingen inderdaad direct en indirect de leeropbrengsten van peer evaluatie beïnvloeden. Waargenomen leeropbrengsten worden voorspeld door zowel concepties over evalueren als unanimiteit in doelen. De concepties worden vervolgens weer voorspeld door psychologische veiligheid, vertrouwen en unanimiteit in doelen.

De uitkomsten van deze studies leren ons dat psychologische veiligheid en unanimiteit in doelen niet alleen hoger zijn in een peer evaluatieomgeving dan in een docentgestuurde evaluatieomgeving; deze variabelen hebben ook een positieve invloed op de leeropbrengsten. Een conclusie kan dan zijn dat waargenomen leeropbrengsten hoger zullen zijn alleen al door een peer evaluatieomgeving te

implementeren die aan bepaalde kenmerken voldoet, en studenten actief betrekt bij het proces van evalueren.

### 2.3 Een nieuwe setting: 360 graden feedback

In de laatste empirische studie van dit proefschrift onderzoeken we de rol van interpersoonlijke opvattingen van lerenden in een werkomgeving; namelijk binnen de specifieke peer evaluatieomgeving van 360 graden feedback. Het doel van deze studie is om te onderzoeken of de rol van interpersoonlijke opvattingen zoals we die hebben gevonden in schoolomgeving ook binnen een werkomgeving een rol spelen. De resultaten laten zien dat de relatie tussen de gemeten interpersoonlijke opvattingen en de waargenomen prestaties als resultaat van 360 graden feedback wordt verstrekt door het vertrouwen dat de lerende in zichzelf en in de collega als evaluator heeft. Concepties, transparantie en de interpersoonlijke opvattingen psychologische veiligheid, unanimiteit in doelen en onderlinge afhankelijkheid voorspellen vertrouwen in zichzelf en de collega als evaluatoren.

## 3 Tot slot

In het algemeen kunnen we zeggen dat de resultaten van de studies het theoretische kader dat is gepresenteerd in de inleiding van dit proefschrift onderbouwen: het belang van interpersoonlijke opvattingen in peer evaluatieomgevingen wordt onderstreept. Voor de dagelijkse praktijk betekent dit voor docenten dat, om optimale winst te halen uit peer evaluatie, de evaluatieomgeving goed doordacht en opgezet moet zijn: het concept peer evaluatie wordt duidelijk uitgelegd, onderlinge interactiemogelijkheden worden besproken met de studenten en evaluatiecriteria worden gezamenlijk geformuleerd. Alleen dan zullen de studenten de interpersoonlijke context positiever ervaren, en daardoor zullen de waargenomen leeropbrengsten toenemen. Een vergelijkbaar advies kan worden gegeven aan managers die 360 graden feedback willen implementeren: de sleutel tot succes van het implementeren van deze vorm van evalueren zit in het betrekken van de medewerkers bij het proces van het opzetten en uitvoeren ervan.

In het zoeken naar mogelijkheden voor toekomstig onderzoek komen we in hoofdstuk 6 met een aantal ideeën, waarvan we er hier twee bespreken. Dit proefschrift laat zien dat psychologische veiligheid en unanimiteit in doelen hoger zijn in een peer evaluatieomgeving dan in een docentgestuurde evaluatieomgeving. Gezien het feit dat iedere evaluatieomgeving anders is, en de dynamiek in elke groep anders is, verwachten we dat het proces van creëren van psychologische veiligheid en unanimiteit in doelen moet plaatsvinden steeds wanneer studenten participeren in een nieuwe peer evaluatiesituatie, omdat de peergroup telkens verschillend

kan zijn. Echter, eerdere ervaring met peer evaluatie zou dit proces wellicht makkelijker of sneller kunnen laten verlopen bijvoorbeeld omdat studenten meer vertrouwd zijn met de uitgangspunten, organisatie en procedures van peer evaluatie. In dit licht zou toekomstig onderzoek waardevolle kennis kunnen toevoegen.

Ten tweede geeft deze dissertatie resultaten weer van waargenomen effecten zoals gerapporteerd door de studenten of werknemers. Omdat in het onderzoek geen relatie is gevonden tussen waargenomen leeropbrengst en de leeropbrengst die de docent en de peers aangeven, is het de vraag of de prestatiematen die we in tot nog toe in onze onderzoeken hebben gebruikt voldoende sensitief zijn om de complexiteit van leeropbrengsten van peer evaluatie te omvatten. Toekomstig onderzoek kan hier meer diepte aan geven door te investeren in het ontwerpen van meet sensitieve beoordelingsmaten, en die te koppelen aan de interpersoonlijke variabelen.

## Referenties

Assessment Reform Group (2002). *Assessment for learning: 10 principles*. Cambridge, University of Cambridge School of Education.

Birenbaum, M., & Dochy, F. J. R. C. (1996). *Alternatives in assessment of achievements, learning processes and prior knowledge.* Boston, Dordrecht, London: Kluwer Academic Publishers.

Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*, 350-381.

Lingard, L., Reznick, R., Espin, S., Regeher, G., & Devito, I. (2002). Team communications in the operating room: Talk patterns, site of tension and implications for novices. *Academic Novices, 77*, 232-237.

Segers, M., Dochy, F., & Cascallar, E. (2003). *Optimising new modes of assessment: In search of qualities and standards.* Dordrecht, Boston, London: Kluwer Academic Publishers.

Van den Bossche, P., Gijselaers, 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 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.

# Appendix

Appendix 1
*Criteria form*

| Do all components work? | |
|---|---|
| **Component** | **Score** |
| Electric motor works | Good / Poor |
| Hydraulic valve works to order | Good / Poor |
| Pneumatic valve works to order | Good / Poor |
| Bascom A (programming device) works | Good / Poor |
| Bascom B works | Good / Poor |
| Electronic component (emitter, receiver, converter) | Good / Poor |
| Scoring for the purpose of safety | Good / Poor |
| Details and finishing | |
| Cables neatly concealed | Good / Poor |
| Finishing (painted, polished, etc.) | Good / Poor |
| Originality | Good / Poor |
| Works according to design | Good / Poor |