



Universiteit
Leiden
The Netherlands

Assessing together : Peer assessment from an interpersonal perspective

Gennip, A.E. van

Citation

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

Version: Not Applicable (or Unknown)

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

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

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

Cover Page



Universiteit Leiden



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

Author: Gennip, Anna Eva (Nanine) van

Title: Assessing together. Peer assessment from an interpersonal perspective

Issue Date: 2012-10-23

CHAPTER 2

Peer Assessment for Learning from a Social Perspective: The Influence of Interpersonal Beliefs and Structural Features⁵

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 & William, 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, Gijsselaers, 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, Gijssels 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? 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 or supplementary?
7 Official weight (how)	Contributing to assessee's 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/assessee's?
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?

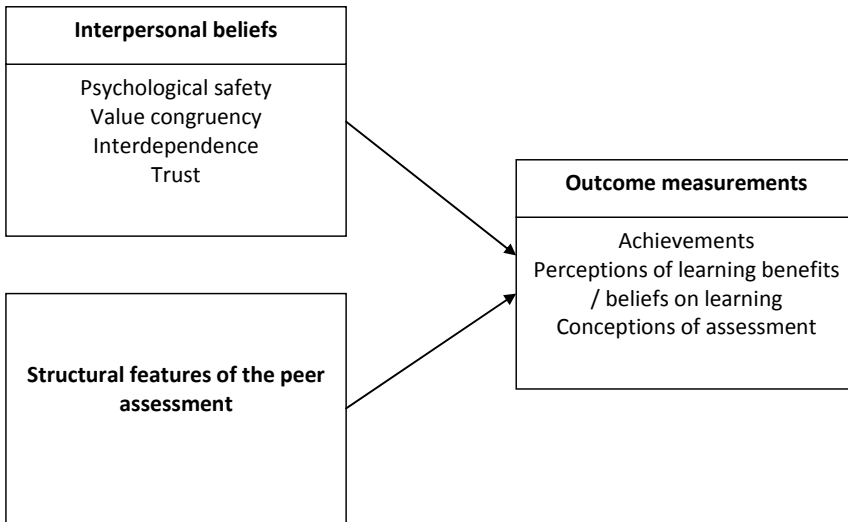


Figure 1
Structural model of analysis

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
Althaus, 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	-
Sluifsmans, D. M. A., Brand-Gruwel, S., & Merriënboer, 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, L., 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
Althaus, R., & Damall, K. (2001)	Mutual, confidential, distance	Same ability, assessors and assessed individual	-	-	-	-
Sluijsmans, D. M. A., Brand-Gruwel, S., & Merriënboer, 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), Althausen & 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), Althausen & 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 Althausser 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 (Althausser & 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 assesseses. 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

- Althausen, 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., & William, 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.

Chapter 2

- 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. Li. (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., Gijssels, 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., Gijssels, 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.

